## Sécurité des Systèmes d'information et des Réseaux

**Raphaël Marichez**
**ENST Paris**
**Mars 2006**

# Les méthodes de lutte contre le spam

## Recherche systématique de méthodes par analyse chronologique de l'acheminement du courrier électronique.


R. Marichez – raphael.marichez@polytechnique.org – http://falco.bz

# 0. Préambule

Ce mémo suppose que le serveur de mail considéré prend déjà en compte les “Best Practices” de la RFC 2505 (et particulièrement de la section 2.), telles que les restrictions usuelles sur l'utilisation de la fonction de relai, ou l'acceptation des Delivery Status Notifications (DSN, ou “Bounces”). Il suppose également une connaissance basique des commandes les plus standards du protocole SMTP (RFC 2821).

La quantité de spam se mesure comme dans la théorie du signal, par analogie avec le rapport signal-sur-bruit, donc en proportion, ou en décibels, et non en nombre de mails.
La mesure pertinente est donc de savoir si on **divise** le spam par 2 ou par 10, non qu'on le diminue de 50%.
Un filtre va éliminer 80% des spams, puis un second 95%, puis un troisième 50%.
On préfera dire que la diminution est d'un facteur 5, puis d'un facteur 20, puis d'un facteur 2. Le total est facile à calculer , 5*20*2 = 200.
Cependant, il nous faut également considérer le risque d'erreur des filtres, qui augmente souvent avec leur sensibilité et donc leur efficacité. N'oublions pas enfin que certains tratements ont des coûts qui peuvent être notables.

Les critères des différentes méthodes présentées sont notés de 0 à 5, par ordre croissant d'efficacité, de pereinité, de risque de faux-positifs, ou de gravité lors du traitement d'un faux-positif. On appelle un faux-positif, un message abusivement classifié comme spam.
On évaluera arbitrairement l'efficacité globale du filtre ainsi :

- Effet faux-positifs : FP = FP.N (nombre) * (FP.I/5) (impact)
- Efficacité globale du filtre = 2*Eff – FP – (Load*Load)/5 – (5-P)

Cette évaluation dépend fortement du profil du MTA considéré. Nous avons essayé d'établir une évaluation moyenne (traffic moyen, 500.000 courriels/jour sur un petit nombre de serveurs grands publics, utilisateurs hétérogènes, mélange de courrier professionnel et personnel, faible médiatisation du nom de domaine, mais présence régulière d'emails sur le web).

## Terminologie :

Conformément aux RFC et aux spécifications des principaux logiciels SMTP (postfix, sendmail), nous appelons :

- **Transaction SMTP** : partie de la session TCP initiée par une commande “MAIL FROM” et terminée par la fin de la commande “DATA”, c'est à dire après avoir envoyé le message SMTP terminé par **<CR><LF>.<CR><LF>**
  Il peut y avoir plusieurs transactions SMTP au sein d'une même session TCP.
  La **transaction SMTP** contient des **commandes SMTP** (EHLO, HELO, MAIL FROM, RCPT TO, DATA, RSET, etc), et au plus un **message SMTP** (initié par commande “DATA”). Ce **message SMTP** est composé d'une série d'**entêtes SMTP** se terminant par une ligne vide, puis d'un **corps SMTP**, se terminant par “**<CR><LF>.<CR><LF>”**
- **Adresse de retour** : l'adresse contenue dans la commande **“MAIL FROM:”** au début de la transaction SMTP, également présente chez le destinataire dans l'entête SMTP **“Return-Path:”**. C'est l'adresse à laquelle doit être adressé un **bounce** en cas de non-livraison du courrier.
- **Entête “From:”** : C'est l'**entête SMTP** contenue dans le **message SMTP**, souvent associée aux entêtes “To:”, “Subject:”, voire “CC:”. Son utilisation n'est pas rigoureusement obligatoire, mais elle permet d'informer l'utilisateur destinataire de l'identité de l'émetteur du courrier, par l'intérmédiare de son logiciel client mail.
- **Destinataire** (ou **destinataire réel** dans les cas d'ambiguïté) : adresse mail spécifiée par la ou les commandes **RCPT TO:** , qui peuvent ne pas figurer dans les **entêtes SMTP** “To:” et “Cc:” (cas des copies conformes invisibles).
- **Bounce**: ou **DSN**, **Delivery Status Notification**, message d'erreur ou d'avertissement envoyé à l'adresse de retour en cas de non-livraison du courrier, définitive (erreur 500, over-quota, user unknown, ...) ou temporaire (timeout, serveur failure, too busy, DNS error, ...).
- **Faux-positif** : un courrier légitime injustement classifié en spam.
- **Faux-négatif** : un vrai spam non classifié en spam.
- **Serveur SMTP** : service écoutant sur le port 25 d'une adresse réseau et destinée à recevoir du courrier.
- **Client SMTP** : client TCP destiné à contacter des serveurs SMTP afin de délivrer du courrier.
- **MX** : acronyme pour Mail eXchanger, utilisé ici pour parler des serveurs destinés à recevoir le courrier pour un domaine “domain.tld”, et figurant dans les enregistrements MX du DNS de “domain.tld”.
- **MTA** : acronyme pour Mail Transport Agent, qui peut être un MX, mais également un client SMTP destiné à délivrer sur internet du courrier émanant de l'intérieur d'un réseau privé.
- **Client mail** : l'interface utilisateur pour lire et écrire du courrier : Outlook (Express), (Mozilla) Thunderbird, Mail, Kmail, Eudora, mutt, pine, elm, mailx, ou un webmail (Squirrelmail, Yahoo, gmail...).

## 1. Le client smtp cherche à contacter les MX

**Idée** : publier de nombreux enregistrements DNS MX avec des priorités faibles, mais qui ne pointent sur rien. Les spammeurs arrosent les différents MX de manière uniforme.
**Efficacité** : **Eff=4**. pour 90% de MX “fictifs”, le spam diminue d'un facteur 10.
**Faux-positifs** : **FP.N=3**. Environ 10%.
**Impact sur les faux-positifs** : **FP.I=4**. Retard du courrier, s'assimile au greylisting (I=2). Le retard peut être cumulé (I+1). Il n'y a pas de processus d'apprentissage. (I+1)
**Impact de charge** : **Load=0**. Impact uniquement sur l'émetteur
**Utilisation** : jamais observée.
**Contournements** : améliorer le code des mini-clients SMTP des spammeurs.
**Péreinité** : **P=3**. Fonctionne tant que tout le monde ne s'en sert pas.
**Références web** : inconnues.
**Efficacité globale : 3.6**

## 2. Etablissement de la connexion TCP

**Nom : RBL (Real-time BlackLists)**
**Informations disponibles** : IP source, résolution DNS inverse associée, corrélées avec des listes de blacklisting temps-réel "RBL" (mail-abuse.com, rfc-ignorant.org, najbl.org, etc)
**Idée** : refuser ou accepter le paquet TCP SYN selon la présence de l'IP du client dans des listes de blacklisting adaptées. On peut y associer l'absence l'enregistrement DNS inverse.
**Efficacité** : **Eff=2**. Facteur 2 à 4
**Faux-positifs** : **FP.N=1.** Rares et selon les listes choisies. Pour les "residential IPs", il s'agit de particuliers ou de TPE/PME possédant leurs MTA sur des fournisseurs ADSL grand public. Pour "rfc-ignorant", quelques expéditeurs sont légitimes mais non conformes aux RFC.
**Effet sur les faux-posifits** : **FP.I=5.** Létal. Le mail n'est jamais acheminé. L'émetteur est informé.
**Impact de charge** : **Load=1.** Négligeable
**Utilisation** : présente mais peu répandue en France (AOL, polytechnique.fr, ...). Plus présente en Europe (Allemagne...)
**Contournements** : Contournements très difficiles (nécessite la découverte d'open-relays eux même rapidement blacklistés).
**Péreinité** : **P=5.** excellente.
**Références web** : http://www.email-policy.com/Spam-black-lists.htm , http://en.wikipedia.org/wiki/DNSBL , http://rbls.org/
**Efficacité globale : 2.8**

**Nom : Quotas**
**Idée** : Refuser les nouvelles connexions à partir d'un certain quota par minute, par heure ou par jour. On peut aussi limiter le nombre de connexions simultanées. On ajoute souvent à cela un quota sur le nombre de destinataires ("RCPT TO") et/ou le nombre de session SMTP par connection TCP.
**Efficacité** : **Eff=4**. Très efficace contre les bots.
**Faux-positifs** : **FP.N=4.** Très présents (listes de diffusion, "réveil" d'un serveur après une coupure, etc)
**Effet sur les faux-posifits** : **FP.I=4.** Retard du courrier, parfois élevé (une journée).
**Impact de charge** : **Load=1.** Négligeable
**Utilisation** : La plupart des implémentations proposent par défaut des limites très hautes qui ne sont en principe jamais atteintes, sauf en cas de problème de configuration. AOL est connu pour utiliser un système de quota à moyen-terme (quotidien) et mis à jour dynamiquement selon le flux "moyen" de chaque IP. D'autres grands ISP en utilisent également.
**Contournements** : Contournements très difficiles, à moins d'étaler l'émission de spams sur plusieurs jours, ce qui réduit considérablement l'intérêt.
**Péreinité** : **P=5.** excellente.
**Références web** : http://www.postfix.org/rate.html , AOL : http://www.abul.org/article220.html , http://www.abul.org/IMG/pdf/liste-blanche-AOL.pdf
**Efficacité globale : 4.6**

## 3. Début de la transaction SMTP : HELO/EHLO

**Nom : HELO checks**
**Idée** : Refuser la transaction (erreur 500) si le HELO/EHLO n'est pas formaté conformément aux RFC. Certains filtres sont encore plus sensibles (demandent que le nom fourni soit un fully-qualified-domain-name).
**Efficacité** : **Eff=2**. Ne concerne que les clients SMTP codés avec des erreurs, en particulier certains virus qui envoient une commande du type HELO:-5425410
**Faux-positifs** : **FP.N=0.** Jamais observé.
**Effet sur les faux-posifits** : **FP.I=5.** Létal
**Impact de charge** : **Load=1.** Négligeable
**Utilisation** : très fréquente, et parfois très sensibles (ENST par exemple qui demande un fully-qualified-domain-name)
**Contournements** : Il suffit de recoder le client SMTP.
**Péreinité** : **P=2.** Si tout le monde s'en sert, il faudra recoder le client SMTP pour qu'il envoie des HELO corrects.
**Références web** : http://www.postfix.org/spam.html#smtpd_require_helo
**Efficacité globale : 0.8**

**Nom : HELO checks (variante plus sensible)**
**Idée** : Refuser la transaction (erreur 500) si le HELO/EHLO fourni correspond à un domaine inconnu ou blacklisté, ou, mieux, s'il correspond à son propre domaine (ce qui ne devrait pas arriver selon les RFC, même si ce n'est pas une obligation.)
**Efficacité** : **Eff=2**. De nombreux virus ou bots utilisent le HELO correspondant au domaine qu'ils cherchent à contacter (exempe : “HELO m4x.org” pour contacter m4x.org). Cela ne concerne que les virus, mais nous avons observé plusieurs grosses vagues de connexions qui ont été bloquées grâce à cela.
**Faux-positifs** : **FP.N=0.** Jamais observés.
**Effet sur les faux-posifits** : **FP.I=5.** Létal
**Impact de charge** : **Load=1.** Négligeable
**Utilisation** : Vraisemblablement assez rare.
**Contournements** : Il suffit de recoder le client SMTP : le HELO est falsifiable.
**Péreinité** : **P=2.** Si tout le monde s'en sert, il faudra recoder le client SMTP pour qu'il envoie des HELO corrects.
**Références web** : http://www.postfix.org/spam.html#smtpd_helo_restrictions
**Efficacité globale : 0.8**

## 4. Début de la transaction SMTP : MAIL FROM:

**Nom : FQDN check (fully qualified domain name)**
**Idée** : Refuser la transaction (erreur 500) si l'adresse de retour fournie n'est pas complète. Cette mesure a plus un objectif de contrer les systèmes mal configurés que les spams. En effet, il s'agit de mails que nous ne pourront "bouncer" en cas d'erreur (over-quota, user unknown, redirection invalide...)
**Efficacité** : **Eff=1**.
**Faux-positifs** : **FP.N=1.** Il existe des faux-positifs, mais il s'agit de systèmes mal configurés, qu'il convient de prévenir. Certains systèmes refusent malheureusement l'adresse de retour nulle "<>" (DSN, Domain Sender Null), pourtant nécessaire pour leur relayer des bounces. Ces systèmes sont listés chez rfc-ignorant.org dans la catégorie DSN (cf paragraphe suivant).
**Effet sur les faux-posifits** : **FP.I=5.** Létal
**Impact de charge** : **Load=1.** Négligeable
**Utilisation** : Quasiment systématique.
**Contournements** : Il suffit de recoder le client SMTP. L'adresse de retour est falsifiable, les spammeurs ne se soucient pas des bounces.
**Péreinité** : **P=2.** On ne peut pas s'en servir raisonnablement comme mesure anti-spam.
**Références web** : http://www.postfix.org/uce.html#smtpd_sender_restrictions
**Efficacité globale : -2.2**

**Nom : blacklist sur l'adresse de retour ("rhsbl")**
**Idée** : Refuser la transaction (erreur 500) si l'adresse de retour est listée dans une blacklist, qu'il s'agisse d'une blacklist d'émetteurs de spam notoires, ou de domaines mal configurés (rfc-ignorant.org), pour lesquels, souvent, il serait impossible de relayer un bounce
**Efficacité** : **Eff=3**. Efficace sans être exhaustif.
**Faux-positifs** : **FP.N=1.** Il existe des faux-positifs, mais il s'agit souvent de systèmes mal configurés, qu'il convient de prévenir.
**Effet sur les faux-posifits** : **FP.I=5.** Létal
**Impact de charge** : **Load=1.** Négligeable
**Utilisation** : Fréquente.
**Contournements** : Il suffit de recoder le client SMTP. L'adresse de retour est falsifiable, les spammeurs ne se soucient pas des bounces.
**Péreinité** : **P=2.**
**Références web** : http://www.postfix.org/uce.html , http://www.rfc-ignorant.org
**Efficacité globale : 1.8**

**Nom : unlisted sender**
**Idée** : Utilisé lorsque l'adresse de retour est sur un domaine local. Pour être certain que l'adresse de retour soit une adresse valide, certains bots utilisent le domaine destination (exemple : m4x.org) , en y associant un utilisateur arbitraire (john@m4x.org). Lorsque le domaine est local au serveur, nous refusons (erreur 500) le message si l'utilisateur n'est pas connu chez nous.
**Efficacité** : **Eff=2**. Parfois utile.
**Faux-positifs** : **FP.N=1.** Il existe des faux-positifs, mais il s'agit souvent d'utilisateurs ayant mal configuré leur compte mail, qu'il convient de prévenir.
**Effet sur les faux-posifits** : **FP.I=5.** Létal
**Impact de charge** : **Load=1.** Négligeable

**Utilisation** : Courante.
**Contournements** : Si le spammeur ne connait pas à l'avance une adresse valide, le contournement est difficile. Autrement, il est trivial
**Péreinité** : **P=3.**
**Références web** : http://www.postfix.org/postconf.5.html#reject_unlisted_sender
**Efficacité globale : 0.8**

**Nom : unverified sender**
**Idée** : Pour être certain que l'adresse de retour soit valide, le plus sûr est de la tester avec notre propre client SMTP, en stoppant la transaction juste après le RCPT TO: et avant le DATA.
**Efficacité** : **Eff=3**. Efficace, mais certains domaines (yahoo.fr) acceptent tout sans distinction, il est donc impossible, pour eux, de déterminer à l'avance si un utilisateur existe.
**Faux-positifs** : **FP.N=1.** Il existe des faux-positifs, mais il s'agit souvent d'utilisateurs ayant mal configuré leur compte mail, qu'il convient de prévenir.
**Effet sur les faux-posifits** : **FP.I=5.** Létal
**Impact de charge** : **Load=3.** Elevé. En plus des requêtes DNS nécessaires, il faut démarrer une sessions SMTP. L'utilisation d'un cache est nécessaire.
**Utilisation** : Peu fréquente en raison des besoins en ressource. Est bien adapté à un ordinateur personnel ou pour une très petite structure. Peut générer des délais si le serveur en face utilise du greylisting ou le même système.
**Contournements** : Le spammeur doit connaitre à l'avance une adresse valide.
**Péreinité** : **P=3.**
**Références web** : http://www.postfix.org/postconf.5.html#reject_unverified_sender
**Efficacité globale : 2.4**

## SPF – Sender Policy Framework

**Idée** : Le constat de base est qu'il est facile, en SMTP, de forger les adresses d'expédition : adresse d'enveloppe ( “MAIL FROM”, également nommée adresse de retour ou Return-path, c'est l'adresse à contacter pour signifier les problèmes de non-réception grâce aux “bounces), mais aussi l'adresse issue de l'entête “From:” (facultative, contenue dans les entêtes du message SMTP, affichée par le client mail). SPF protège la première de ces deux adresses.

SPF permet d'associer, au domaine de l'adresse de retour, des IP autorisées à émettre du courrier se réclamant de ce domaine. C'est une technique qui s'inspire des domain-keys de Yahoo (voir plus bas). On effectue une requête DNS du type TXT sur le domaine de l'adresse de retour (ou du HELO si l'adresse de retour est nulle dans le cas d'un bounce), et selon le résultat, on considère l'IP du client SMTP comme autorisée, non-autorisée (“hard error” ou “soft error”), ou sans avis (“unknown”).

Contrairement à ce que disent de nombreux vendeurs, SPF est plutôt efficace contre le phishing (usurpation d'un domaine d'une société) que contre le spam de manière générale, pour autant que ce domaine soit bien protégé par SPF, c'est-à-dire que peu d'IP aient le droit d'émettre.

**Efficacité** : **Eff=3**. Si d'une part tout le monde l'utilise, et si d'autre part les utilisateurs itinérants prennent la peine de changer leur adresse d'expédition selon leur situation géographique (ou si les serveurs de courriers sortant utilisent de la réécriture d'adresse comme SRS), alors SPF permettrait de limiter considérablement le nombre d'ordinateurs autorisés à émettre des courriers venant de chaque domaine, réduisant ainsi, potentiellement, considérablement la proportion de spams. SPF ne devient fiable contre le spam que si tout le monde s'en sert, car sinon, il suffit d'utiliser un autre domaine d'émission, peu ou pas protégé par des enregistrements SPF. De plus, comme ce système pose des problèmes pour les organismes qui proposent du relai de messagerie, sa mise en oeuvre est lente.

Actuellement, c'est un système de filtrage à utiliser avec beaucoup de précautions, bien qu'il soit mis en avant par de grands éditeurs de logiciels. SPF reste cependant efficace contre le phishing, mais au prix de faux-positifs possibles.

**Faux-positifs** : **FP.N=4.** Il existe de nombreux faux-positifs, pour essentiellement deux raisons : les organismes qui proposent du relai de messagerie (y compris pour les listes de diffusion n'utilisant pas encore SRS), et les utilisateurs itinérants (notamment, qui envoient du courrier avec une adresse professionnelle depuis leur FAI grand public). SRS (Sender Rewriting Scheme) permet de contourner cette difficulté en ré-écrivant l'adresse de retour avec son propre domaine, de manière transparente pour l'utilisateur, mais il est encore assez peu utilisé. En particulier, il n'est jamais utilisé pour les redirections “simples”, telles que les “.forward” Unix.

Pour ne pas souffrir des effets du SPF, il convient de se rendre compatible SPF, c'est-à-dire de publier des enregistrements DNS SPF, et d'utiliser SRS, mais sans nécessairement utiliser SPF comme filtrage anti-spam.

**Effet sur les faux-posifits** : **FP.I=4.** Létal ou présomption de spam selon le traitement appliqué. Pour un hard-fail, le draft de RFC sur SPF recommande le rejet permanent (erreur 500) du mail (“should”).
**Impact de charge** : **Load=1.** Négligeable (requêtes DNS)
**Utilisation** : SPF est mis en avant par certains grands étideurs. De nombreux grands fournisseurs (dont les adresses sont souvent usurpées) sont protégés par SPF, mais sans nécessairement l'utiliser en filtrage : AOL, Hotmail, Ebay, Hotmail.com, gmx.net, IBM, et d'autres. SpamAssassin v3.0.0 utilise SPF parmi ses critères de filtrage.
**Contournements** : Le spammeur peut modifier son adresse de retour. C'est bien pour cela que le filtre ne fonctionne que si tous les domaines sont protégés par SPF.
**Péreinité** : **P=3.**
**Références web** : http://www.openspf.org/ http://www.ietf.org/internet-drafts/draft-schlitt-spf-classic-02.txt http://en.wikipedia.org/wiki/Sender_Policy_Framework http://www.openspf.org/srs.html
**Efficacité globale : 0.8**

## 5. Début de la transaction SMTP : RCPT TO:

**Nom : Multi-Bounce**
**Idée** : Refuser les “bounces” (Delivery Status Notifications, “MAIL FROM:<>”) lorsqu'ils sont destinées à plusieurs destinataires (“RCPT TO”).
**Efficacité** : **Eff=1**. Si l'utilisation de “MAIL FROM:<>” permet de contourner tous les filtres portant sur l'adresse de retour et vus précédemment, l'envoi d'un unique email à un grand nombre de destinataires a rapidement été bloqué. Ce type de spam n'existe presque pas.
**Faux-positifs** : **FP.N=0.**
**Impact de charge** : **Load=1.**
**Utilisation** : Régulière
**Contournements** : Il suffit simplement de redémarrer une session par un “MAIL FROM” entre chaque mail, au prix d'une charge à peine plus élevée.
**Péreinité** : **P=1.**
**Références web** : http://www.postfix.org/postconf.5.html#reject_multi_recipient_bounce
**Efficacité globale : -2.2**

## Greylisting

**Idée** : Rejeter par défaut tous les mails avec une erreur temporaire (400), juste après la commande “RCPT TO:”, et retenir le triplet {IP-source; expéditeur; destinataire}. On parle de “greylist” comme d'une position intermédiaire entre la blacklist (erreurs 500) et la whitelist (acceptations 200). Tous les triplets non encore connus sont donc par défaut dans la greylist. Lorsqu'un triplet a été tenté (et a généré une erreur 400), il est placé dans la whitelist pour une durée d'au moins un jour.

Un client SMTP légitime va réessayer l'envoi au bout de quelques minutes ou quelques dizaines de minutes. A ce moment, le mail sera accepté car le triplet {IP-source; expéditeur; destinataire} sera trouvé dans la whitelist. La durée d'expiration de ce triplet est alors augmentée à plusieurs jours, voire semaines, pour éviter d'induire un nouveau retard lors des prochaines correspondances similaires.

**Efficacité** : **Eff=5**. Le spam est en gérénal réduit d'un facteur 100 voire 1000, étant donné que les clients SMTP qui véhiculent le spam sont majoritairement de petits clients SMTP embarqués dans des vers sur des ordinateurs infectés, et qui cherchent à innonder au plus vite les serveurs SMTP, sans se soucier de la bonne arrivée à destination des messages.

**Faux-positifs** : **FP.N=5.** Evidemment, tout le monde se trouvant dans la greylist, tout courrier légitime sera retardé par le greylisting, du moins à la première occurrence du triplet {IP-source; expéditeur; destinataire}. Certains clients SMTP légitimes ne réessayent pas après une erreur 400, mais ils sont en général connus et font partie des exceptions.

**Effet sur les faux-posifits** : **FP.I=3.** Retard de courrier, en général relativement acceptable (10 à 60 minutes voire jusqu'à 300), mais dépendant des caractéristiques du client SMTP.

**Impact de charge** : **Load=3.** Elevé et fortement dépendant du trafic de mail. La taille de la whitelist augmente non pas linéairement en fonction du trafic SMTP, mais en carré voire en cube, puisqu'il faut tenir compte des trois éléments du triplet, sachant que l'IP source et l'émetteur sont en général relativement liés. De plus, l'utilisation du greylisting sur des serveurs distribués impose des synchronisations de la whitelist entre les différents MTA, ce qui ajoute à la charge des systèmes.

**Utilisation** : Régulières dans les environnements de recherche ou les petites structures. Impossible ou très coûteuse chez les gros fournisseurs de service mail.
**Contournements** : Difficiles. Les mini-clients SMTP embarqués doivent être recodés pour gérer les codes d'erreur 400 et donc les files d'attente de messages sortant, ce qui est assez difficile, et actuellement inutile compte-tenu de la faible proportion (en poids) des organismes utilisant le greylisting. Nous avons déjà observé quelques rares spams passant au travers du greylisting, mais il s'agit probablement souvent de coïncidences, les délais entre chaque tentatives étant très longs (souvent une journée précise).
**Péreinité** : **P=4.** Fonctionne tant que tout le monde ne s'en sert pas, ce qui serait très coûteux pour les plus gros serveurs de mail.
**Références web** : http://hcpnet.free.fr/milter-greylist/ , http://www.solutionslinux.fr/document_conferencier/43f0d47313fca.pdf
**Efficacité globale : 5.2** En réalité, très efficace (6.8) pour des petites structures où la charge induite est faible, notamment si on tolère un retard régulier de courrier.

## 6. Transaction SMTP : headers SMTP

**Nom : Blacklists sur les entêtes From: , To: , X-Mailer: , Subject: , etc...**
**Idée** : Certaines vagues ponctuelles de spams peuvent être bloquées sur l'identification d'un élément commun aux messages incriminés. Cela nécessite l'intervention d'un administrateur, il est dangereux de se reposer sur du long terme sur ces éléments qui sont aisément falsifiables.
Nous ne pouvons donc pas classer ce type de filtrage dans les filtres automatiques, il est donc délicat de lui attribuer une note comparative.
L'impact de charge est négligeable, mais les faux-positifs sont possibles.
Cette technique a été mise en oeuvre pas de très nombreux administrateurs lors de la dernière grosse vague de mails allemands d'extrême-droite générés par le virus SoberQ, le week-end du 14-15 mai 2005 et les quelques jours suivants. La date de déclenchement de ce type de spams "politiques" (de même que pour le phishing) est souvent un vendredi ou un samedi, de manière à profiter du week-end et de l'absence de certains administrateurs systèmes.
**Références web** : (SoberQ) http://archives.neohapsis.com/archives/linux/mandrake/2005-q2/0077.html

**Nom : BCCs**
**Idée** : Très souvent les spams sont envoyés au sein d'une même transaction SMTP, ne donnant qu'une seule entête SMTP "To:" (voire aucune), aucune entête "Cc:", mais définissant de très nombreux destinataires réels (commande SMTP "RCPT TO:"). Tout se passe comme si nous étions en "BCC" (copie conforme invisible). On peut donc refuser les mails qui contiennent un trop grand nombre de destinataires en BCC, sous certaines conditions vis-à-vis des entêtes "From:" et "To:". Dans une version plus exagérée de ce filtre, justifiée par l'existence de relais qui "découpent" le mail en autant de sessions SMTP que de destinataires réels, nous pouvons aussi classifier comme spam un mail qui contiendrait un seul destinataire en BCC..
**Efficacité** : **Eff=3**. Semble très efficace, mais nous n'avons pas de retour d'expérience dessus.
**Faux-positifs** : **FP.N=3.** Il existe des faux-positifs, mais il s'agit souvent d'utilisateurs effectuant des maillings de masse sur leur carnet d'adresse (ce qui peut être considéré comme du spam), ou bien, plus gênant, de listes de diffusion (il faudra donc les exclure des vérifications).
**Effet sur les faux-posifits** : **FP.I=4.** Létal ou associé à la politique anti-spam globale.
**Impact de charge** : **Load=2.** Faible, mais demande d'attendre la délivrance du courrier et le traitement des entêtes.
**Utilisation** : Nous savons que Hotmail/MSN utilise ce système, malheureusement de manière exagérée : une présomption de spam accrue est assignée à un mail dont l'unique destinataire réel diffère de l'entête "To:", ce qui est le cas des listes de diffusion, des redirections de courrier, et également des personnes recevant du courrier légitime en étant en BCC.
**Contournements** : Le spammeur devrait ajouter les destinataires réels aux entêtes SMTP "To:" ou "Cc:", ce qui complique sa tache, mais est possible. Comme tous les destinataires seront facilement visibles, le spammeur devra alors, pour rester discret, limiter drastiquement le nombre de destinataires par session SMTP, ce qui générera plus de flux, car plus de sessions SMTP. Notons que ce système viendrait combler l'une des limitations signalées des Domains Keys, puisqu'alors les entêtes SMTP changent à chaque mail, et il faut alors recalculer la signature pour chaque destinataire.
**Péreinité** : **P=3.**
**Références web** : aucune. (internes Microsoft).
**Efficacité globale : 0.8**

## Yahoo! Domain Keys

**Idée** : Un peu comme pour SPF, on essaye d'empêcher l'envoi d'emails semblant venir de son propre domaine, mais envoyés par des personnes non légitimes. Pour empêcher l'usurpation d'email (email spoofing), un mécanisme de signature à clef publique est utilisé dans le champs d'entêtes “DomainKey-Signature:”. La clef publique est disponible dans le DNS du domaine, et les serveurs de courrier sortants sont chargés de signer le message entier et d'ajouter la signature aux entêtes SMTP.
Le but est de relier le domaine émetteur tel qu'indiqué dans le client mail (entête “From:user@domain.tld”) au serveur ayant relayé le courrier. Contrairement à SPF, qui permet d'assurer que le client SMTP est bien autorisé par le domaine décrit dans l'adresse de retour (Return-path), les Domain Keys assurent que le serveur prétendu par le domaine de l'entête “From:” a effectivement validé et transmis lui-même le message. C'est-à-dire qu'il a accepté le message interne venant de l'utilisateur, l'a signé, puis l'a transmis sur internet. Ce mécanisme suppose donc la bonne configuration du serveur de courrier sortant en terme de restrictions anti-spam vis-à-vis des utilisateurs internes, ce qui n'est malheureusement que rarement le cas.

Exemple : Si nous recevons un mail venant de “From:user@domain.tld”, nous allons vérifier à l'aide de la signature contenue dans l'entête “DomainKey-Signature:”, et à l'aide de la clef publique du domaine “domain.tld”, que ce mail a bien été signé par le serveur de courrier sortant de “domain.tld”.
**Efficacité** : **Eff=3**. Par conséquent, de la même manière que SPF, ceci ne peut fonctionner en tant que mesure anti-spam que si tout le monde l'utilise, puisqu'il suffirait alors de choisir un domaine non protégé par les Domain Keys. De plus, ceci n'empêcherait pas un spammeur d'acheter son propre nom de domaine, et d'émettre du spam qu'il aurait lui-même signé. Par contre, en tant que mesure de lutte contre le phishing, ce mécanisme est bien efficace, au prix de faux-positifs possibles.
**Faux-positifs** : **FP.N=4.** Toujours de la même manière que SPF, il existe des faux-positifs, au travers des relais de messagerie, et des utilisateurs itinérants. Cependant, SRS ne serait ici d'aucune utilité, puisque l'entête “From:” est visible de l'utilisateur contrairement à l'adresse de retour. Une modification de l'entête “From:”, outre le fait d'être désagréable pour l'utilisateur destinataire, casserait les signatures PGP ou S-MIME.
**Effet sur les faux-posifits** : **FP.I=3.** Fait partie d'autres critères anti-spam. Le draft de RFC relatif aux Domain Keys propose (“may”) le rejet permanent (erreur 500) de messages non signés ou de signature incorrecte, lorsque le domaine est protégé. A titre de comparaison, le draft de RFC sur SPF écrit “should”.
**Impact de charge** : **Load=2.** Faible (requêtes DNS), mais demande d'attendre la délivrance du courrier (après la commande DATA, le corps SMTP suit immédiatement les entêtes SMTP, il n'est pas autorisé d'interrompre la connexion).
**Utilisation** : Encore peu répandue, si ce n'est (évidemment) par Yahoo, qui utilise également SRS pour ses listes de diffusion, mais n'a pas publié d'enregistrements SPF.
**Contournements** : Comme SPF, il suffit au spammeur de modifier son adresse d'expédition “From:”. Il faudrait donc que tout le monde utilise ce système pour qu'il soit utilisable. Cependant, même dans ce cas, rien n'empêche un serveur malveillant de signer des spams. On peut même imaginer que des mini-clients SMTP embarqués sur les ordinateurs infectés envoient du spam préalablement signé par des Domain Keys, puisque le message SMTP n'a pas besoin de changer (l'entête de destination “To:” est totalement indépendante du destinataire réel “RCPT TO:”). Un tel domaine serait rapidement connu comme émetteur de spam, mais ceci nous ramène à l'utilisation de méthodes de blacklist classiques évoquées plus haut. Nous considérons donc les contournements des Domain Keys plus faciles que ceux de SPF.
**Péreinité** : **P=2.**
**Références web** : http://en.wikipedia.org/wiki/Domain_keys http://antispam.yahoo.com/domainkeys http://tools.ietf.org/wg/dkim/draft-ietf-dkim-base/draft-ietf-dkim-base-00.txt
**Efficacité globale : -0.2**

## 7. Transaction SMTP : contenu SMTP

**Nom : Filtrage par contenu statique (URL, etc)**
**Idée** : Souvent (mais pas toujours), le spam est utilisé pour inviter les internautes à cliquer sur des liens web (URL), que ce soit à but de phishing ou de vente, ou bien pour transmettre des messages politiques (car de SoberQ en mai 2005). On peut alors refuser les messages contenant des URL ou des phrases connues pour se trouver dans les plus gros spams du moment.
**Efficacité** : **Eff=4**. Très efficace, car s'attaque aux spams les plus fréquents. Mais cela demande des interventions manuelles des administrateurs pour mettre à jour la base de données. (Nous n'avons pas connaissance de l'existence d'aucune liste temps-réel interrogeable par DNS, comme pour les RBL.) Cette méthode est donc réservée aux serveurs administrés par des personnes présentes en tout temps, donc dans les grosses entreprises.
**Faux-positifs** : **FP.N=1.** Les faux positifs ne seraient que des forwards ou des réponses aux spams eux-mêmes, donc rarement du courrier important.
**Effet sur les faux-posifits** : **FP.I=4.** Létal ou associé à la politique anti-spam globale.
**Impact de charge** : **Load=3.** Impact sensible : il faut traiter l'ensemble du message SMTP avec un processeur de texte ou d'expressions régulières.
**Utilisation** : AOL.
**Contournements** : Presque impossibles, puisqu'on s'attaque à l'objectif même du spammeur, au-delà du simple envoi d'un courrier. Le spammeur pourrait utiliser des extensions javascript, mais elles ne sont pas toujours activées sur les clients mails, et il est également aisé de détecter un tel code. Pour les messages politiques sans URL, une image pourrait remplacer le texte.
**Péreinité** : **P=4.**
**Références web** : aucune. (internes AOL).
**Efficacité globale : 4.4**

**Nom : Filtrage statistique (bayésien)**
**Idée** : Détecter d'après un traitement du contenu du message, souvent en le découpant en mots, la présence de mots souvent trouvés dans les spams, ou au contraire, dans les mails légitimes. Attribuer une probabilité de présence dans un spam à chacun de ces mots, puis aggréger le résultat pour attribuer la note finale au mail. Les entêtes SMTP sont incluses dans ce traitement. Cela nécessite un apprentissage préalable en fournissant au filtre des spams et des mails légitimes.
**Efficacité** : **Eff=4**. Très efficace, pourvu que la base de données soit bien représentative. Elle est d'autant plus représentative que l'échantillon des utilisateurs est fin : même langue, même milieu socio-professionnel. Le cas idéal serait un filtre individuel, mais l'apprentissage est plus long et est souvent fastidieux. Nous sommes en train d'étudier la possibilité de combiner une base de données collectives, à des bases de données individuelles, plus petites, permettant ainsi d'affiner les détections selon chacun.
**Faux-positifs** : **FP.N=1.** Là encore, la présence de faux-positifs dépend directement de la qualité de la base de données. L'irruption d'une langue rare pour la communauté d'utilisateurs considérée (comme le russe ou le chinois, langues souvent repérées dans des spams, ou l'allemand après l'attaque de SoberQ) pose problème. Il convient de fournir aux utilisateurs une possibilité de renseigner le filtre avec des faux-positifs ainsi que des faux-négatifs : on ne peut donc pas supprimer les courriers (sauf ceux avec les plus hauts scores), mais il faut les marquer (dans une entête ou dans le sujet) ou les classer (pour le cas des serveurs IMAP) pour que l'utilisateur puisse y avoir finalement accès si nécessaire. Par voie de conséquence, cela limite l'intérêt du filtre, puisque le message arrive *in fine* dans la boîte de l'utilisateur, bien que classé dans un autre dossier.
Nous renvoyons aux transparents d'une conférence que nous avons tenue en février 2006 aux Solutions Linux pour des discussions plus précises : http://falco.bz/docs/Presentation_bogo.pdf

**Effet sur les faux-posifits** : **FP.I=4.** Létal ou associé à la politique anti-spam globale.
**Impact de charge** : **Load=3.** Impact sensible : il faut traiter l'ensemble du message SMTP avec un processeur de texte avancé.
**Utilisation** : Chez l'utilisateur final (Thunderbird, Apple Mail, parfois SpamAssassin), ou dans des communautés à faible effectifs (SpamAssassin, très gourmand en ressources). Dspam ou bogofilter sont utilisés plus rarement (Polytechnique.org, ...)
**Contournements** : Comme précédemment, le contournement est très difficile, puisqu'il faut modifier drastiquement le contenu du message. Les modifs de remplacements (tel que “viagra” qui devient “v|agra”) sont très vites pris en compte par ce type de filtre, puisque ces mots n'apparraissent jamais dans des messages légitimes. Cependant, l'utilisation de plus en plus fréquente d'HTML (voire des CSS), et, pire, des images, rend ce filtrage moins efficace (il reste tout de même les entêtes SMTP).
**Péreinité** : **P=4.**
**Références web** : Logociels dspam, bogofilter, ou SpamAssassin (qui combine plusieurs techniques). Bogofilter : http://falco.bz/docs/Presentation_bogo.pdf Théorie de Bayes et implémentations : :http://www.paulgraham.com/spam.html , http://www.linuxjournal.com/article.php?sid=6467
**Efficacité globale : 4.4**

**Nom : Filtrage combiné (SpamAssassin)**
**Idée** : Effectuer un traitement des entêtes et des messages, en attribuant une note pour chacun des critères : notamment, un filtrage statistique, un filtrage sur le contenu (URL), mais également des caractéristiques propres aux spams (entêtes HTML avec la présence de couleurs, de gras, de grandes polices, mais également la présence d'images ou d'autres pièces jointes).
Une note globale est attribuée en additionnant les différentes notes obtenues (qui peuvent être positives ou négatives) avec des pondérations paramétrables. Le filtre bayésien peut optionnellement être automatiquement instruit grâce à la classification issue des autres critères, même si cela peut nuire à la qualité de la base de données (instabilité).
**Efficacité** : **Eff=4**. Très efficace, faisant appel à une combinaisons de filtrages, et permettant d'éviter à l'utilisateur débutant le fastidieux travail d'apprentissage initial du filtre bayésien.
**Faux-positifs** : **FP.N=2.** Des messages légitimes comportant les caractéristiques de spams selon SpamAssassin (couleur dans le HTML, etc) sont parfois classifiés spams. Il conviendrait d'attacher de l'attention lors de l'apprentissage du filtre bayésien, puis ensuite d'associer une plus forte pondération au le filtrage bayésien qu'au filtrage HTML.
**Effet sur les faux-posifits** : **FP.I=4.** Létal ou associé à la politique anti-spam globale.
**Impact de charge** : **Load=5.** SpamAssassin est très gourmand en ressources : le processeur de texte n'est pas linéaire contrairement aux autres filtrages de contenu, en particulier à cause du filtrage HTML.
**Utilisation** : Chez l'utilisateur final, ou dans des communautés à faible effectifs.
**Contournements** : Comme précédemment, le contournement est très difficile, puisqu'il faut modifier drastiquement le contenu du message.
**Péreinité** : **P=4.**
**Références web** : Logociels dspam, bogofilter, ou SpamAssassin (qui combine plusieurs techniques). Bogofilter : http://falco.bz/docs/Presentation_bogo.pdf Théorie de Bayes et implémentations : :http://www.paulgraham.com/spam.html , http://www.linuxjournal.com/article.php?sid=6467
**Efficacité globale : 0.4**

## Discussion :

Dans une transaction SMTP, tout est falsifiable sauf :
- l'adresse IP de l'émetteur
- la finalité du spam, donc son contenu.

Ce qui explique les notes élevées des filtrages suivants :
**Quotas** (AOL)
**Greylisting** (petites structures, et lorsque le retard n'est pas pénalisant)
**Filtrage statique** (Besoin d'administrateurs sept jours sur sept) (AOL)
**Filtrage statistique bayésien** (communautés homogènes)

Il faut cependant garder à l'esprit que ce système de notation est fortement dépendant de l'organisme considéré : par exemple, SpamAssassin est très efficace chez l'utilisateur, mais bien trop gourmand en ressources pour être utilisé à large échelle. De même, un filtrage bayésien est efficace au sein d'une communauté d'utilisateurs homogène, mais moins pour un grand domaine internationnal comme aol.com .

De manière générale, c'est l'association de plusieurs techniques qui va permettre à un organisme d'être réellement protégé contre le spam. Par exemple, considérons les **blacklists RBL** et le **greylisting**, deux techniques relativement efficaces. Cependant, prises séparément :
- Les RBL génèrent des faux positifs permanents en la personne des particuliers qui ont leur propre client SMTP, ou des PME/TPE qui sont sur des fournisseurs d'accès grands publics et utilisent leurs propres MTA. Les faux-positifs sont en petit nombre mais l'impact est majeur.
  **RBL : moyennement efficace (facteur 4) ; peu de faux-positifs.**
- Le greylisting génère toujours des faux positifs, même si l'impact est faible (retard de courrier uniquement).
  **Greylisting : très efficace (facteur 1000); impact mineur sur les faux-positifs.**

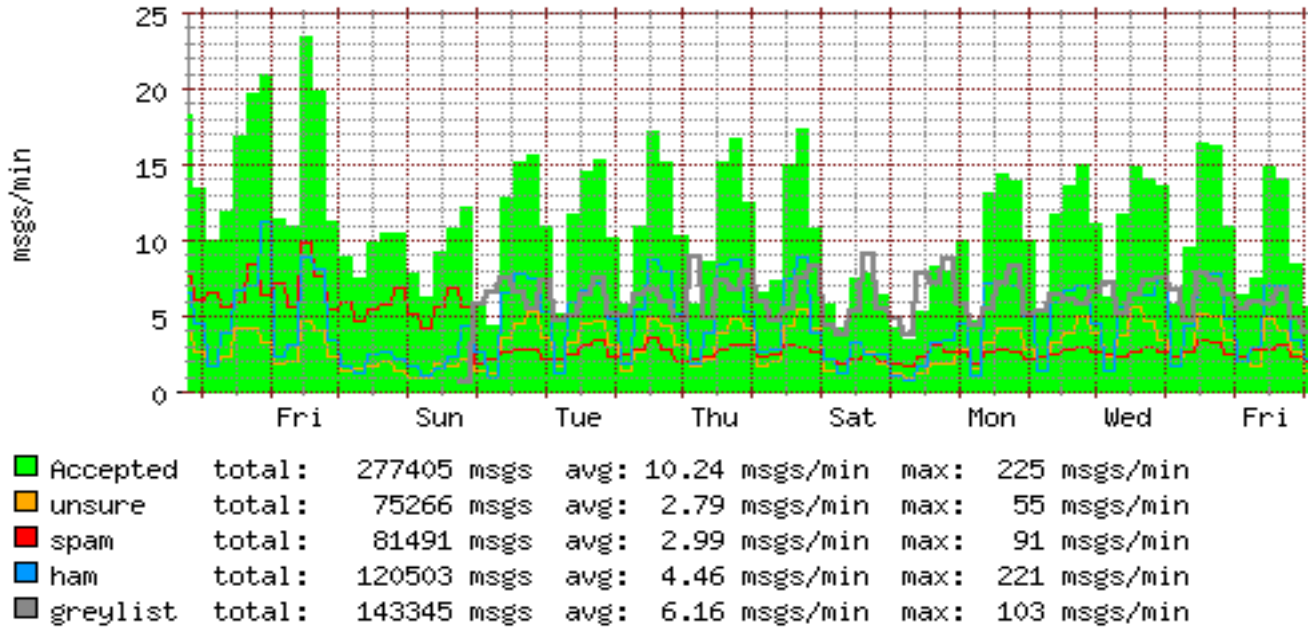
En associant intelligemment ces deux filtres, pour conjuguer leurs actions sur les faux-positifs, nous arrivons à des résultats vraiment bons. Pour arriver à cela, un développeur Debian de l'association Polytechnique.org (madcoder@debian.org) a codé le logiciel **whitelister**.
Après les contrôles usuels et peu coûteux (HELO, FQDN, rfc-ignorant, unlisted_sender), le démon whitelister va accepter le mail si l'expéditeur n'est pas listé dans les RBL, ou transmettre le mail au greylisting **uniquement si** l'expéditeur est listé. On a ainsi peu de faux positifs (RBL), et, quand bien même il y en aurait, l'impact est faible (retard de courrier).
Soyons honnête, l'efficacité est bien plus faible qu'avec un greylisting total, car nous filtrons alors les mails qui sont sensibles aux **deux** filtres à la fois, et le RBL n'est pas très efficace.

Utilisons alors, en dernier lieu, un filtre très efficace, peu coûteux, mais susceptible de quelques faux-positifs. Comme la proportion de spams a déjà été divisée par environ 4, nous pouvons nous permettre de laisser passer les spams, mais en les marquant, pour que l'utilisateur soit certain de ne rien perdre. Le candidat tout choisi est donc le filtre bayésien.

Ici, nous serions tentés d'effectuer le même traitement avec le filtre bayésien : transmettre au greylisting les mails qui seraient détectés comme spams. C'est malheureusement impossible actuellement, car le greylisting doit avoir lieu **avant** la fin de la commande SMTP **DATA**. Or nous sommes obligés d'accepter de prendre en charge le mail pour effectuer le traitement bayésien. Rigoureusement parlant, les normes nous autorisent à répondre une erreur 400 à la fin de la commande DATA, mais il faudrait pour cela que notre filtre bayésien nous donne sa réponse avant d'avoir accepté

définitivement le mail, ce qui n'est pas actuellement le cas : nous utilisons Postfix qui attend d'avoir reçu le <CR><LF>.<CR><LF> et d'avoir répondu une acceptation (200) avant qu'il ne puisse transmettre le message à bogofilter.
Les filtres les plus légers en terme de charge induite sont ceux qui nous permettent de refuser le message avant la commande DATA. Pour ceux-ci, ce refus peut prendre la forme d'une erreur temporaire 400 pour réduire les impacts sur les faux positifs, même si cette solution (hormis avec le greylisting) n'existe pas sur le marché. Seul le couple RBL-greylisting tire parti de cette possibilité aujourd'hui.

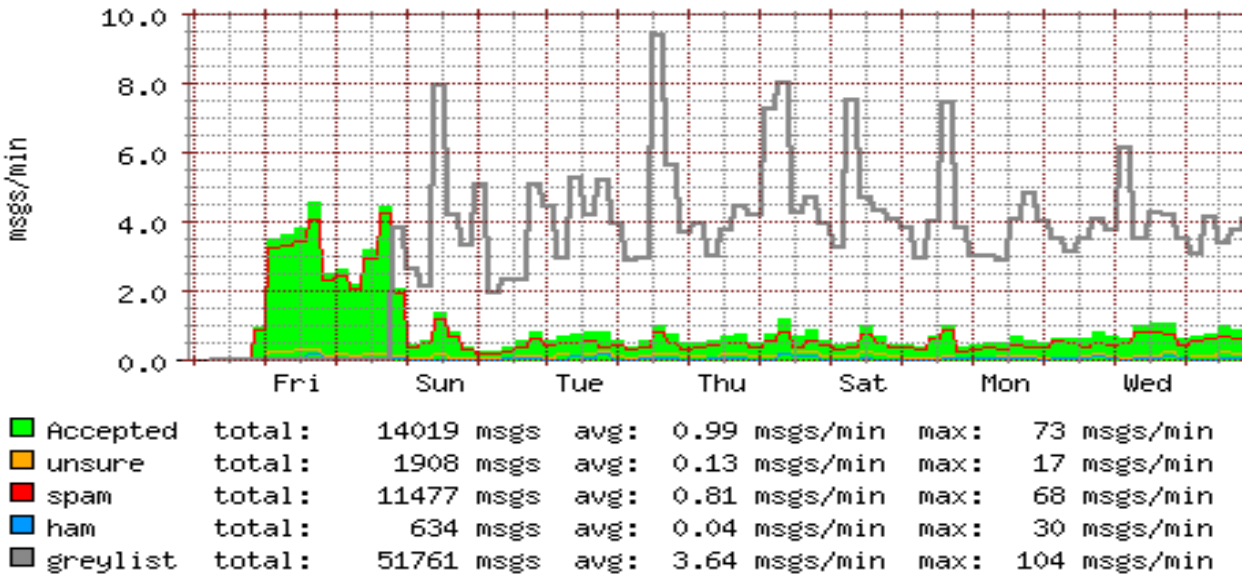
Sur l'un des deux serveurs principaux de **Polytechnique.org** (13 000 membres, 300 000 courriers par jour), nous avons mis en place cette association il y a quelques semaines. Le filtre bayésien vient après, avec l'anti-virus.




La surface verte représente les mails acceptés par notre MTA (après les contrôles usuels, puis l'anti-virus). La ligne rouge est la quantité de messages détectés comme spams par notre filtre bayésien, bogofilter. La ligne bleue est la quantité de messages détectés comme messages légitimes par notre filtre bayésien.

Lorsque nous avons mis en place le greylisting, la quantité de spams (détectés par bogofilter) a été divisée par 3 à 4, comme prévu. Les mails résistants au greylisting sont de l'ordre de 20 par semaine sur l'ensemble de nos serveurs, et parmi ceux-ci, la majorité sont eux-mêmes des spams. Pour ce qui est de nos 13 000 utilisateurs, l'impact sur leurs mails légitimes est donc négligeable (au pire, du retard). Par contre, l'impact sur la taille de leur dossier “spams” est notable.

Sur notre serveur de secours, qui est listé dans les enregistrements MX de notre DNS, mais avec une priorité plus faible, l'action du couple RBL-greylisting est spectaculaire :




Comme ce serveur ne transmet presque que du spam (la courbe rouge se confond presque avec la surface verte), l'action du couple RBL-greylisting a été excellente. Elle a même dépassé nos attentes, ce qui parait logique, puisque les expéditeurs les plus sujets à envoyer des mails à notre serveur de secours sont ceux qui ne gèrent pas les priorités des enregistrements MX, donc les mini-clients SMTP embarqués sur les ordinateurs compromis par des vers ou des virus... qui sont pour la plupart chez des particuliers, donc pour la plupart listés dans les RBL.
Globalement, la quantité de mails marqués “spams” par notre filtre bayésien a donc été divisée par cinq à six. La quasi-totalité des spams restants sont ceux passant via nos deux serveurs principaux, et qui, de plus, proviennent de serveurs SMTP très légitimes, relayant des listes de diffusions ou des redirections de mail. Ces relais ne sont donc pas sujets au test RBL-greylisting, même si l'émetteur initial, avant le relai, aurait dû être bloqué par ce test s'il avait eu lieu sur le relai et non chez nous.

Avant de mettre en place une politique anti-spam, il convient donc de choisir, parmi les solutions existantes, une combinaison qui corresponde à la fois au profil des utilisateurs, à leurs usages, et également aux ressources systèmes et humaines disponibles. Aucune méthode unique ne peut résoudre le problème en apportant simultanément une forte efficacité et l'absence de faux-positifs.