Lutte Anti-Spam à l'aide d'outils libres

Matthieu Herrb

20 janvier 2005

CNRS-LAAS

Rappels

- SPAM (pollupostage, pourriels,...) : couriers électroniques non sollicités,
- Collecte de listes d'adresses sur le Web, Usenet, etc... puis revente ad nauseam,
- Activité plus ou moins illégale mais lucrative,
- Abus de relais SMTP ouverts ou de machines piratées pour masquer l'identité du spammeur

Vers la fin du mail ?

- Utilisateurs exaspérés
- De plus en plus de messages perdus indirectement à cause du SPAM :
  - faux positifs
  - destruction accidentelle
- Perte de confiance dans l'outil
- Retour au papier ?

Techniques de lutte

- Au début : bouton 'poubelle' $\longrightarrow$ la méthode la plus simple, taux d'erreur très bas.
- Listes noires (utilisation du DNS)
- Filtrage du contenu :
  - par mots clés,
  - bayesien.
- Greylisting
- Identification forte de l'expéditeur : SPF, Sender-Id, DomainKeys...
- Pots de miel.

Les trois lois de l'anti-spam

- **Ne pas altérer les messages valides**
  (signature électronique)

- **Respecter la vie privée et la législation**
  (Rapports de la CNIL, jurisprudence...)

- **Ne pas ajouter de bruit dans le système**
  (Messages de rejet, erreurs incompréhensibles, délais exagérés etc.)

L'administrateur système et la pollution électronique

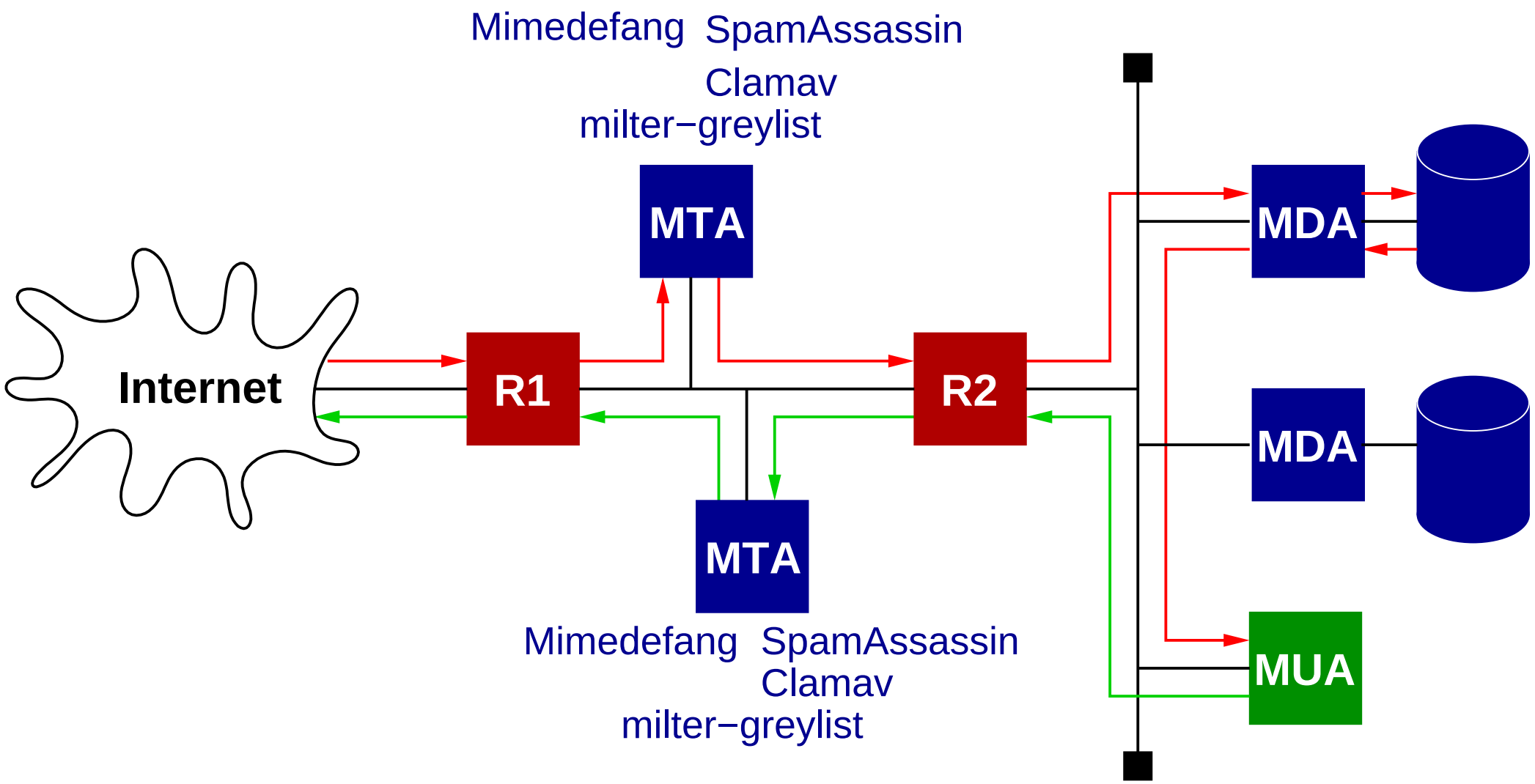
- (In)former ses utilisateurs
  - Cf. recommandations de la CNIL
  - Ne pas devenir spammeurs (organisation de conférences, gestion de listes de diffusion, etc.)
- Mettre en place un dispositif anti-spam au niveau du serveur de messagerie
- Lutter contre les relais ouverts.
- Dénoncer à la justice les pratiques illégales (pédophilie, « chaînes », etc.)

Anti-SPAM / Anti-virus au LAAS

CNRS-LAAS : Laboratoire du département STIC du CNRS
- 650 utilisateurs
- 13000 messages entrants / jour avant anti-spam (estimation)
- serveur de listes sympa.

Solution mise en place :
- 2 Sun Netra X1 (Solaris 8) : SMTP entrant + sortant
- Serveurs de Bal séparés
- Sendmail
- Mimedefang
- SpamAssassin
- File : :Scan + Clamav (antivirus)
- milter-greylist

Architecture de la messagerie au LAAS



Listes noires

Le LAAS a utilisé des black lists de 1997 à janvier 2003.

Principe : Rejeter les connexions SMTP en provenance de relais ouverts connus.

Fonctionnement : Un serveur DNS gère un domaine spécial contenant une base de donnée de relais ouverts.

Exemple : Est-ce que `140.93.0.15` (`mail.laas.fr`) est un relais ouvert ?

$\longrightarrow$ recherche `15.0.93.140.relays.ordb.org`.

Si réponse, alors c'est un relais ouvert...

Configuration : (sendmail/m4) Dans le fichier `.mc` ajouter :

```
FEATURE(‘dnsbl’,‘relays.ordb.org’,‘"550 rejected - see http://ordb.org/"’)
```

Possibilité d'exceptions dans `accessdb` : `140.93.0.15 OK`

Quelques listes noires :

rbl.net	payant
SpamCop.net	payant
http://ordb.org/	
http://spews.org/	

Problèmes des listes noires

- Contenu non maîtrisé
- Plus facile d'y rentrer que d'en sortir
- Souvent trop agressives (bloque toute une classe C pour un relais ouvert)
- problème récent Osirusoft : s'est mis à bloquer tout l'internet.
- etc.

Restent intéressantes en complément d'autres outils d'évaluation.

Listes blanches

Complémentaire extrême des blacks lists :

N'accepter que les messages de personnes identifiées

- despam : http://www.laas.fr/~felix/despam.html
- http://impressive.net/people/gerald/2000/12/spam-filtering.html
- n'accepter que des messages signés (S/MIME ou PGP) avec ou non une liste d'autorités de certification de confiance.

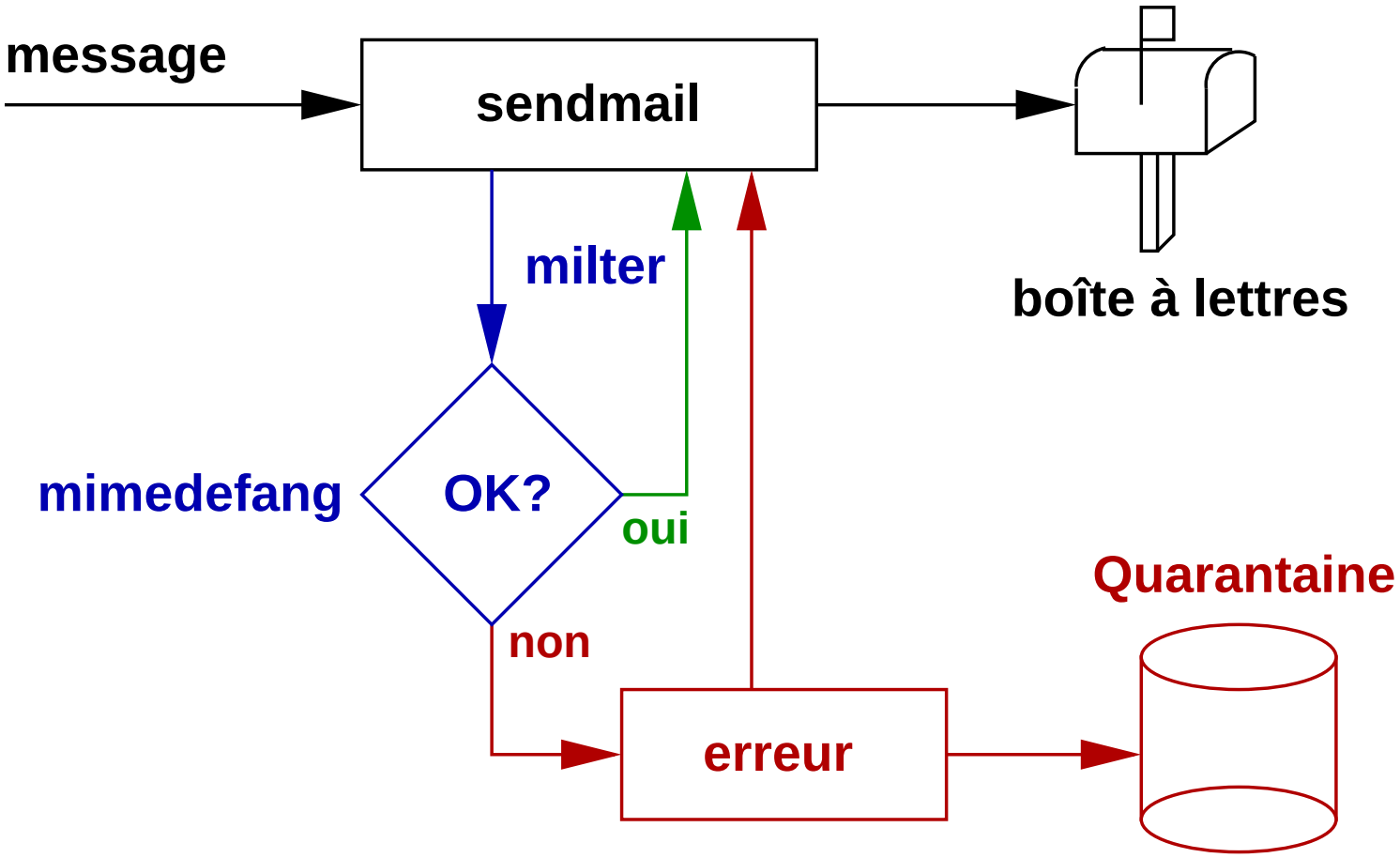
Mimedefang

http://www.roaringpenguin.com/mimedefang/
http://www.mimedefang.org/

Programme général de filtrage de mails, utilisant l'API milter de Sendmail.
- destruction, modification ou mise en quarantaine de fichiers attachés « dangereux »
- interface avec Anti-virus existants, avec SpamAssassin
- ajout de notices aux messages
- actions configurables en fonction du domaine, de l'utilisateur, du relais utilisé, etc.
- utilise un « pool » de processus pour les serveurs chargés.
- ...

$\longrightarrow$ Permet de faire anti-virus et anti-spam en même temps.

Interaction sendmail/mimedefang



Mimedefang - filtres

`mimedefang-filter` est un script Perl.

Utiliser SpamAssassin :

Dans `filter_end` :

```
# Spam checks if SpamAssassin is installed
if ($Features{"SpamAssassin"}) {
    if (-s "./INPUTMSG" < 100*1024) {
        # Only scan messages smaller than 100kB.
        my($hits, $req, $names, $report) = spam_assassin_check();
        if ($hits >= $req) {
            action_change_header("X-Spam-Score", "$hits $names");
            action_change_header("X-Spam-Status", "Yes");
        }
    }
}
```

Mimedefang - exemple

Alias internes seulement

```
sub filter_recipient {
    my ($recipient, $sender, $ip, $hostname, $first, $helo) = @_;

    # Block mail from outside to some aliases
    if ($recipient =~ /<(ace|all|ejm|ita|mac|pl|ria|team|tsf)\@(\w+\.)?laas\.fr>/ &&
        ( $sender !~ /laas\.fr/ || $ip ne "127.0.0.1" && $ip !~ "140\.93\.")) {
          md_syslog('notice',
            "from=$sender, to=$recipient, relay=$name [$ip]: non local sender");
          return ('REJECT', "User unknown", "550", "5.5.1");
      }
}
```

Rejette les messages vers ces adresses (par ex. `all@laas.fr`) si le relais SMTP n'est pas une machine du réseau du LAAS.

Filtrage sur le contenu : principe

Classifier les messages en fonction du texte complet du message
- Par mots clés / patterns : SpamAssassin
- Par analyse statistique : Classification Bayesienne (bogofilter, bmf, etc.)
- Base de donnée de spams : Vipul's Razor.

Autres critères de filtrage

Rejeter les messages mal formés :
- respect des RFC
- entêtes incohérents (User-Agent forgé par ex.)
- caractères illégaux

$\longrightarrow$ rejette SPAM et virus/vers naïfs.

SpamAssassin™

http://spamassassin.apache.org/

Ensemble de tests sur le contenu
Chaque test attribue des points
Somme des points → score

Marque les messages qui dépassent un seuil.

Écrit en Perl.

Plusieurs modes de fonctionnement :
- filtre simple (utilisation avec procmail)
- filtre client d'un démon (meilleurs performances) (`spamc`)
- au travers de l'API milter (sendmail) → filtrage global

SpamAssassin - Exemple

```
SPAM: -------------------- Start SpamAssassin results ---------------------
SPAM: This mail is probably spam.  The original message has been altered
SPAM: so you can recognise or block similar unwanted mail in future.
SPAM: See http://spamassassin.org/tag/ for more details.
SPAM:
SPAM: Content analysis details:   (8.10 hits, 5 required)
SPAM: SUBJECT_MONTH      (-0.5 points) Subject contains a month name - probable newslet
SPAM: NO_REAL_NAME       (1.3 points)  From: does not include a real name
SPAM: LOSE_POUNDS        (0.5 points)  Subject talks about losing pounds
SPAM: DIET               (0.4 points)  BODY: Lose Weight Spam
SPAM: FULL_REFUND        (0.4 points)  BODY: Offers a full refund
SPAM: CLICK_BELOW        (0.3 points)  BODY: Asks you to click below
SPAM: SPAM_PHRASE_08_13  (1.4 points)  BODY: Spam phrases score is 08 to 13 (medium)
SPAM:                    [score: 8]
SPAM: DATE_IN_FUTURE_06_12 (1.1 points)  Date: is 6 to 12 hours after Received: date
SPAM: RCVD_IN_DSBL       (3.2 points)  RBL: Received via a relay in list.dsbl.org
SPAM:                    [RBL check: found 251.102.96.210.list.dsbl.org]
SPAM:
SPAM: -------------------- End of SpamAssassin results ---------------------
```

Tests - SpamAssassin

Sur une base de messages triée à la main.

SpamAssassin :

nature	total	erreurs	pourcentage
HAM	1129	1	0.1%
SPAM	633	22	3.5%

Quelques secondes par message.

Mimedefang + SpamAssassin pour les utilisateurs

Chaque message soupçonné par par SpamAssassin d'être un SPAM est marqué par des entêtes :

```
X-Spam-Score: 7.9 (*******) ALL_CAP_PORN,CLICK_BELOW,CLICK_BELOW_CAPS,
  CTYPE_JUST_HTML,DATE_MISSING,FROM_ENDS_IN_NUMS,LINES_OF_YELLING,
  LINES_OF_YELLING_2,LINES_OF_YELLING_3,PORN_4,SPAM_PHRASE_05_08,
  SUPERLONG_LINE,UPPERCASE_50_75
X-Spam-Status: Yes
X-Scanned-By: MIMEDefang 2.28 (www . roaringpenguin . com / mimedefang)
```

Filtrage par chaque utilisateur avec :

- procmail (http://www.procmail.org/)
- logiciel de messagerie : Eudora, Netscape messenger, ...

Mimedefang + SpamAssassin - problèmes

Besoin d'aide aux utilisateurs pour configurer les MUA !

Dans cette configuration de SpamAssassin :
- pas de personnalisation des filtres / pas de filtre bayésien
- pas d'analyse des entêtes (messages passés par un relais ouvert)
- seuil fixé à 5.0

Problèmes de charge de la machine

Nombre croissant de SPAM non détecté.

Classificateurs bayesiens

Classification bayesienne :
http://www.mathpages.com/home/kmath267.htm

Eric Horvitz & al. (Microsoft) 1998 :
http://research.microsoft.com/~horvitz/junkfilter.htm

Paul Graham 2002-2003 :
http://www.paulgraham.com/spam.html

Adaptive latent semantic analysis (Apple Mail.app) :
http://lsa.colorado.edu/papers/dp1.LSAintro.pdf

Implémentations libres :
http://bogofilter.sourceforge.net/ (Eric S. Raymond)
http://www.sourceforge.net/bmf/
http://spambayes.sourceforge.net/
http://www.mozilla.org/mailnews/spam.html (Mozilla Mail)
http://www.fourmilab.ch/annoyance-filter/

Rappels

$$P(C = c_k|X = x) = \frac{P(X = x|C = c_k)P(C = c_k)}{P(X = x)}$$

Ici deux classes : C_0 HAM
C_1 SPAM

Bayes naïf :

$$P(X = x|C = C_k) = \prod_i P(X_i = x_i|C = c_k)$$

Horvitz & al.

Brevet Microsoft (1998).

Classificateur Bayes naïf.

Utilise une liste de motifs discriminants déterminée par apprentissage. Supprime les mots trop peu fréquents pour être discriminants.

Prise en compte d'éléments spécifiques : proportion de caractères non alphanumériques, domaine de l'expéditeur.

Paul Graham

Décomposition en tokens du texte complet (y compris en-têtes)

Base d'apprentissage classée à la main. Pour chaque token calcule la probabilité d'être SPAM.

Nouveaux messages :
- assigne initialement 0.4 comme probabilité aux mots inconnus ,
- garde les 15 probabilités les plus significatives (les plus éloignées de 0.5),
- calcule la probabilité conditionnelle que le mail soit un SPAM,
- en fonction de la décision recalcule les probabilités des tokens.

Implémentations :
- bogofilter
- bmf

Classification par score ou probabiliste ?

Qu'y a-t-il derrière un score ? Comment les attribuer ?

Scores $\longrightarrow$ difficiles à personnaliser

Probabilités $\longrightarrow$ valeurs absolues

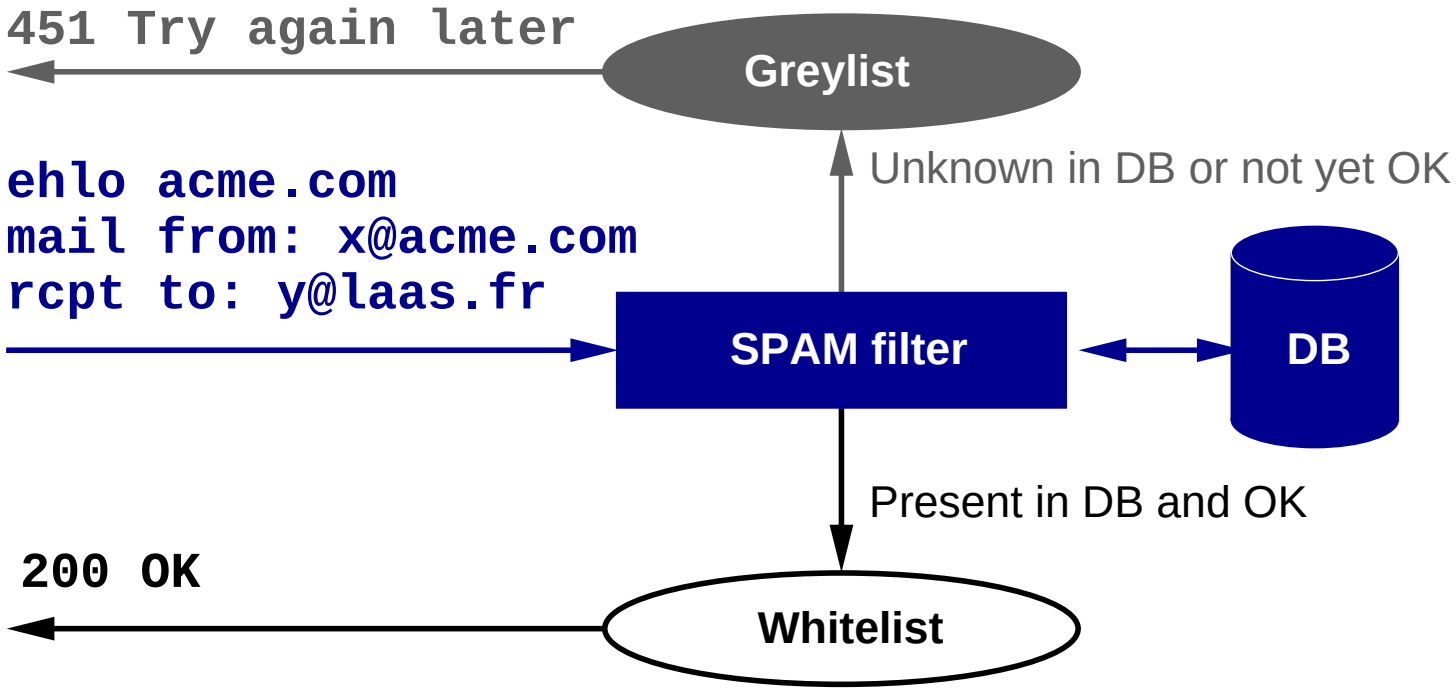
Apprentissage personnalisé automatique (2 boutons : SPAM/HAM)

Permet de suivre l'évolution des contenus du SPAM.

Greylisting

- Technique proposée par
  http://projects.puremagic.com/greylisting/
  (Quelques idées équivalentes apparues en même temps)
- Principe :
  - les spammeurs n'ont pas les moyens de traiter les ré-émissions
  - **refuser** avec un code 4xx (erreur temporaire) les triplets (expéditeur, destinataire, relais) inconnus. Enregister le triplet dans la liste « grise ».
  - Si nouvelle soumission du même triplet dans un certain délai → accepter et mettre en liste blanche.
  - Si pas ne nouvelle soumission dans ce délai : probablement du SPAM.

Greylisting : principe



Greylisting : en pratique

Nombreuses implémentations
- implémentation en Perl de puremagic
- intégré dans spamd (OpenBSD)
- milter-greylist par E. Dreyfus (http://hcpnet.free.fr/milter-greylist)
- greylistd (exim http://packages.debian.org/unstable/mail/greylistd)
- gld (postfix http://www.gasmi.net/gld.html)

Réduit la charge sur le serveur en limitant le nombre de messages à traiter.

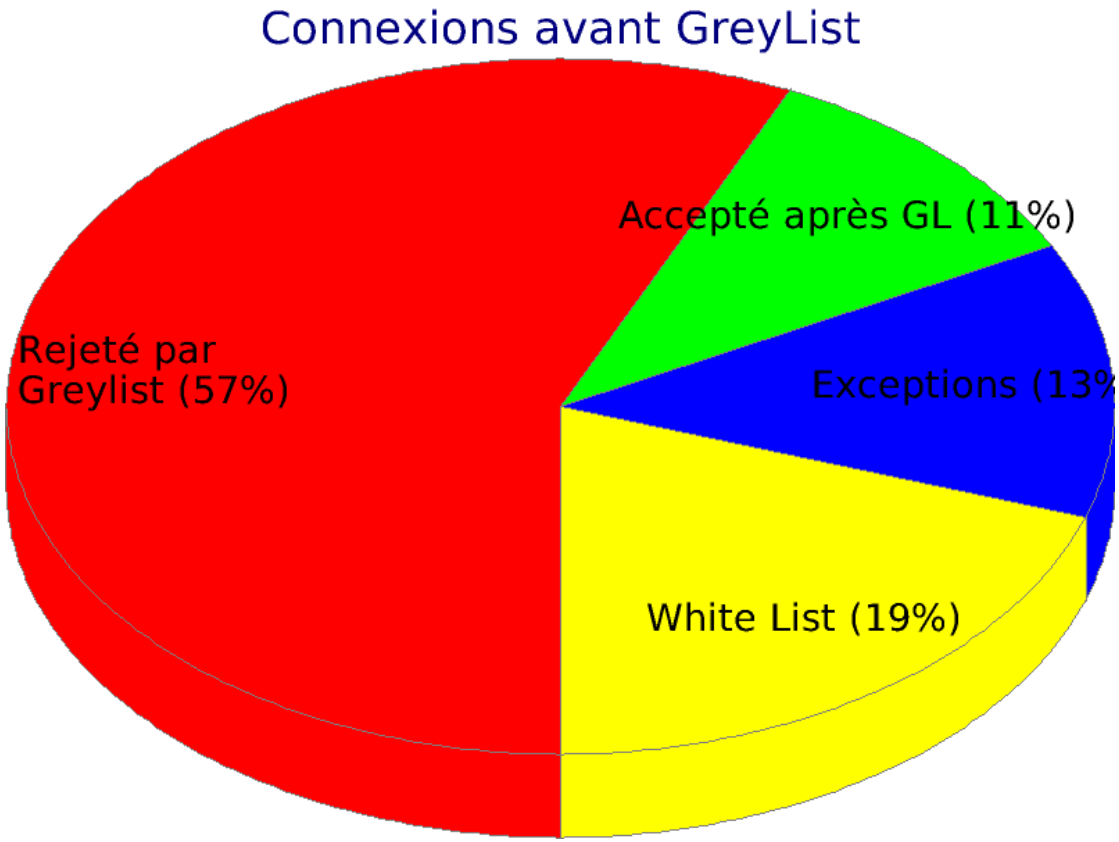
Paramètres utilisés au LAAS :
- Durée de rejet initiale : 25 mn
- Délai pour accepter un message : 3 jours
- Durée de maintien en liste blanche sans activité : 5 semaines

Greylisting : quelques problèmes...

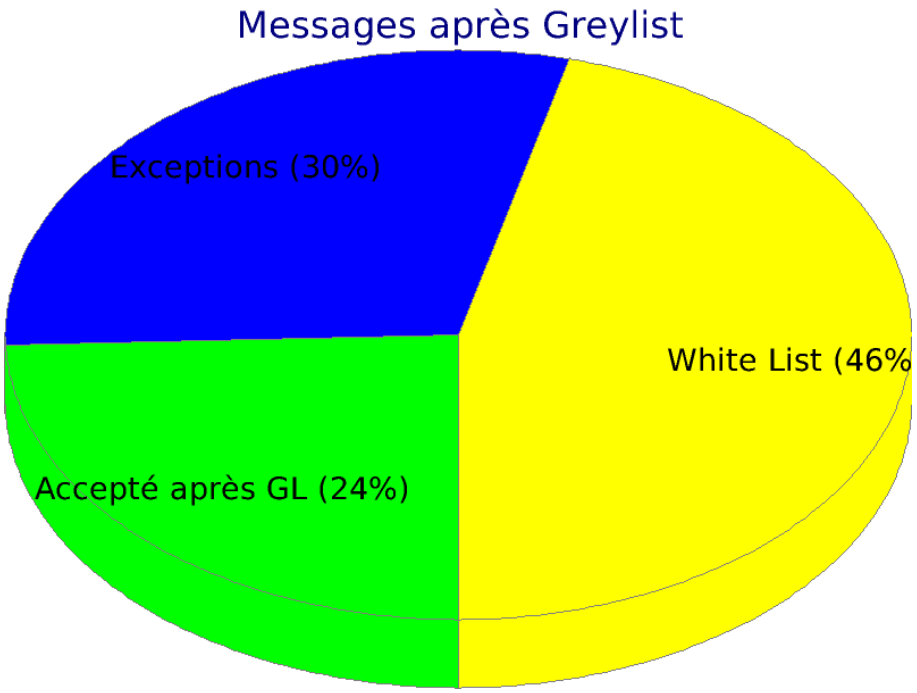
- Gestion des MX : installer un filtre sur chaque serveur.
- La redirection des messages « blanchi » les spams.
- Certaines listes de diffusion qui utilisent un expéditeur fictif unique pour chaque message : jamais de white-list.
- Certains sites (Yahoo groups entre autre) ne gèrent pas les erreurs 4xx : pertes de messages.
- Certains sites remontent directement les erreurs 4xx aux MUA.

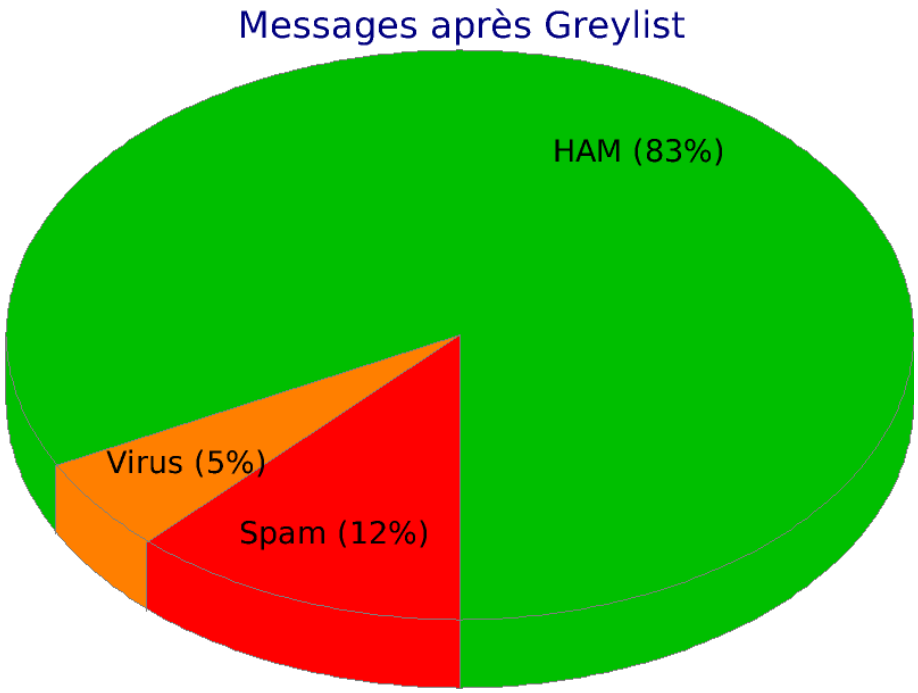
$\longrightarrow$ gestion de listes d'exceptions (liste blanche).

Résultats du greylisting

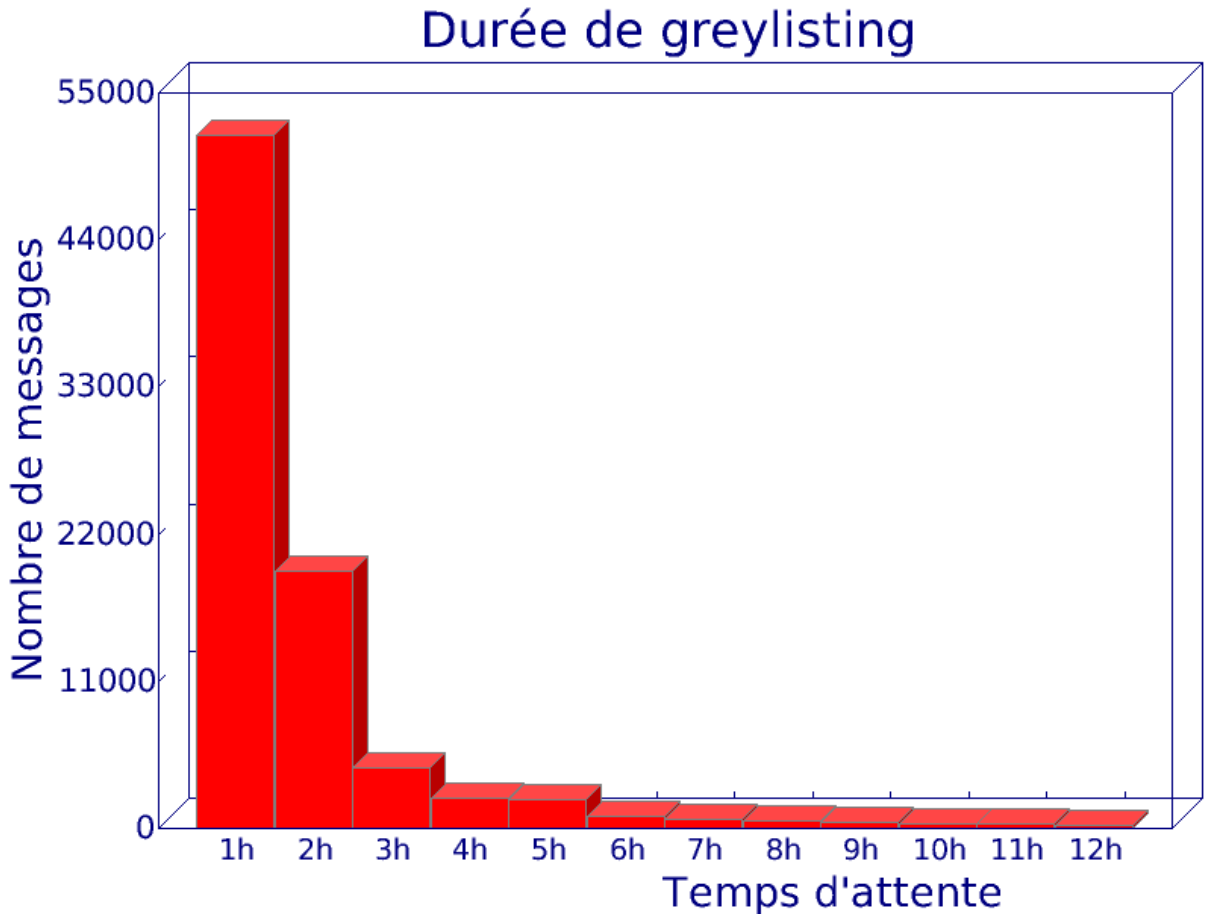


Répartition des messages acceptés





Durée d'attente



Anti-anti SPAM

- Variations sur les mots clés : V|agra
- Attaque en saturation sur les bases d'apprentissage Bayesiennes
- Utilisation de botnets plutot que de relais ouverts
  - plus grand nombre de machines
  - possibilités d'attaques contre sites anti-spam
- Exploitation massive de botnets pour contourner les greylists ?
- De plus en plus de SPAM légal (ie respectant la loi des États-Unis)
  - expéditeur non caché
  - lien pour se désabonner
  - contenu peu masqué
- SPAM plus civilisé ?
  - succès des lois / actions en justice / mesures dissuasives ?
  - diminution du nombre de vulnérabilités exploitables pour diffuser du spam de façon anonyme ?

Sur les contenus exécutables

Problème beaucoup plus vaste que la messagerie...

Technologie en plein boum : web services, mises à jour automatiques d'OS ou de logiciels, ...

Problème de sécurité majeur : vecteurs de diffusion de virus/chevaux de troie idéal (plus besoin de débordement de buffer pour exécuter du code malicieux

Les « bacs à sable » (Java, .NET, ...) n'arrivent pas à être imperméables

La signature numérique des contenus exécutables est mal partie

Bibliographie

- R. Costales & E. Allman **Sendmail, 3rd edition**, O'Reilly.
- R. Blum, **Postfix**, SAMS.
- P. Graham - A plan for spam
  http://www.paulgraham.com/spam.html
- Misc 17 - Comment lutter contre le spam, les malwares, les spywares ?
- http://spam.abuse.net/
- http://www.spam.com/ le site web de Hormel.