La lutte anti-SPAM. Exemple avec DSPAM

Fabrice Prigent

Université Toulouse 1 Sciences Sociales

Jeudi 7 février 2008

Une situation qui se dégrade

Selon les chiffres de Postini

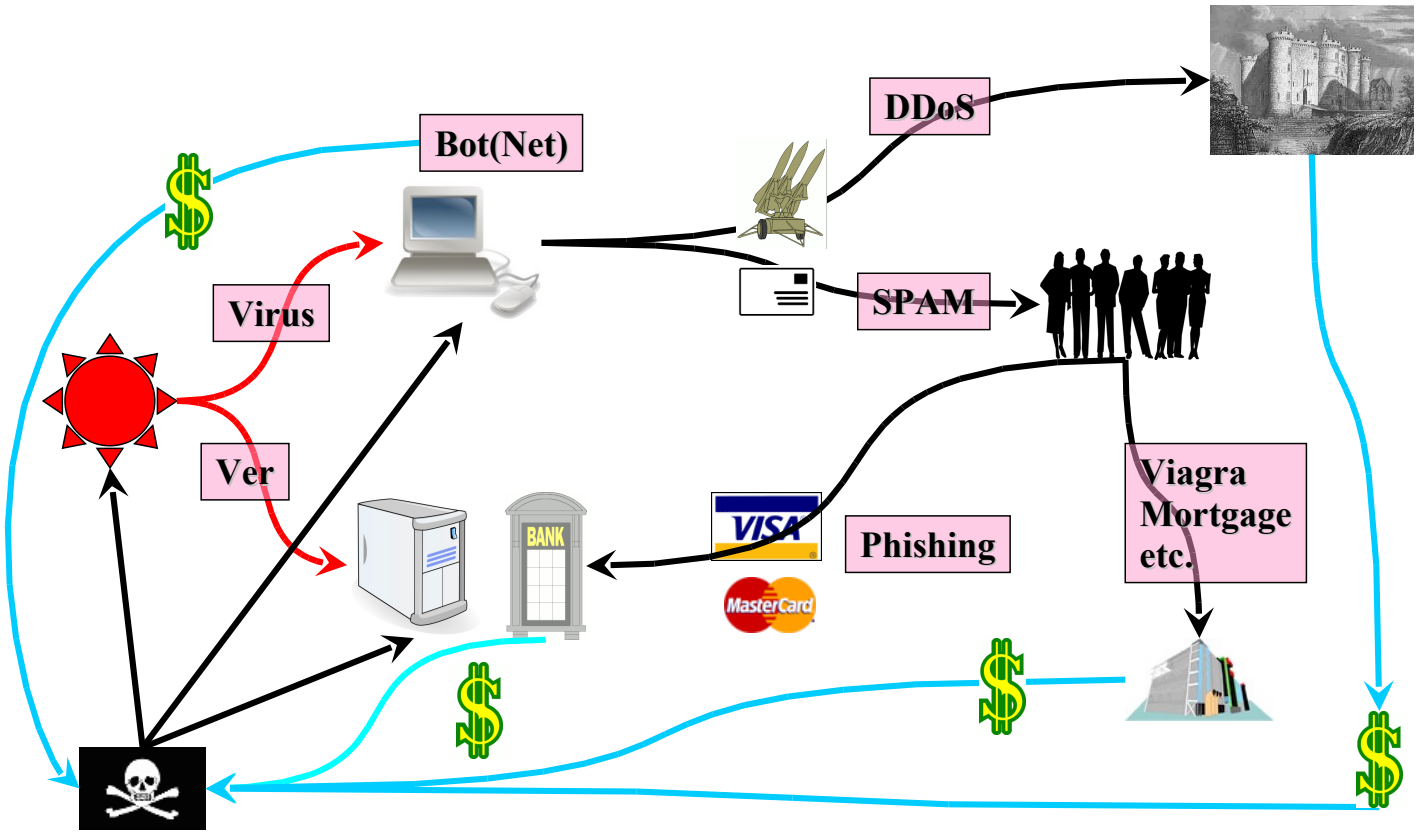
- 95% de spams en 2006
- 99% en 2007

Des spammeurs plus performants

La présence des ingénieurs chez les spammeurs porte ses fruits

- Utilisation massive des zombies (RBL souvent inutiles),
- Les zombies renvoient les courriers (greylisting moins efficace),
- Les spams images complexifient la détection (utilisation d'outils d'O.C.R.),
- Insertion de textes aléatoires,
- Contournement grâce à l'utilisateur (enlever '*' dans l'url, ou url sans point www misssado com)

Un écosystème qui a fait ses preuves



Pour résumer

- Pirate = Virus = Spam = Scam = Phishing = Argent

Quelques chiffres du SPAM

Les revenus annexes

- 572 dollars : coût moyen d'un vol d'identité bancaire
- 80 000 dollars : pour une faille
- Rock Phish : 200 Millions de dollars de C.A.

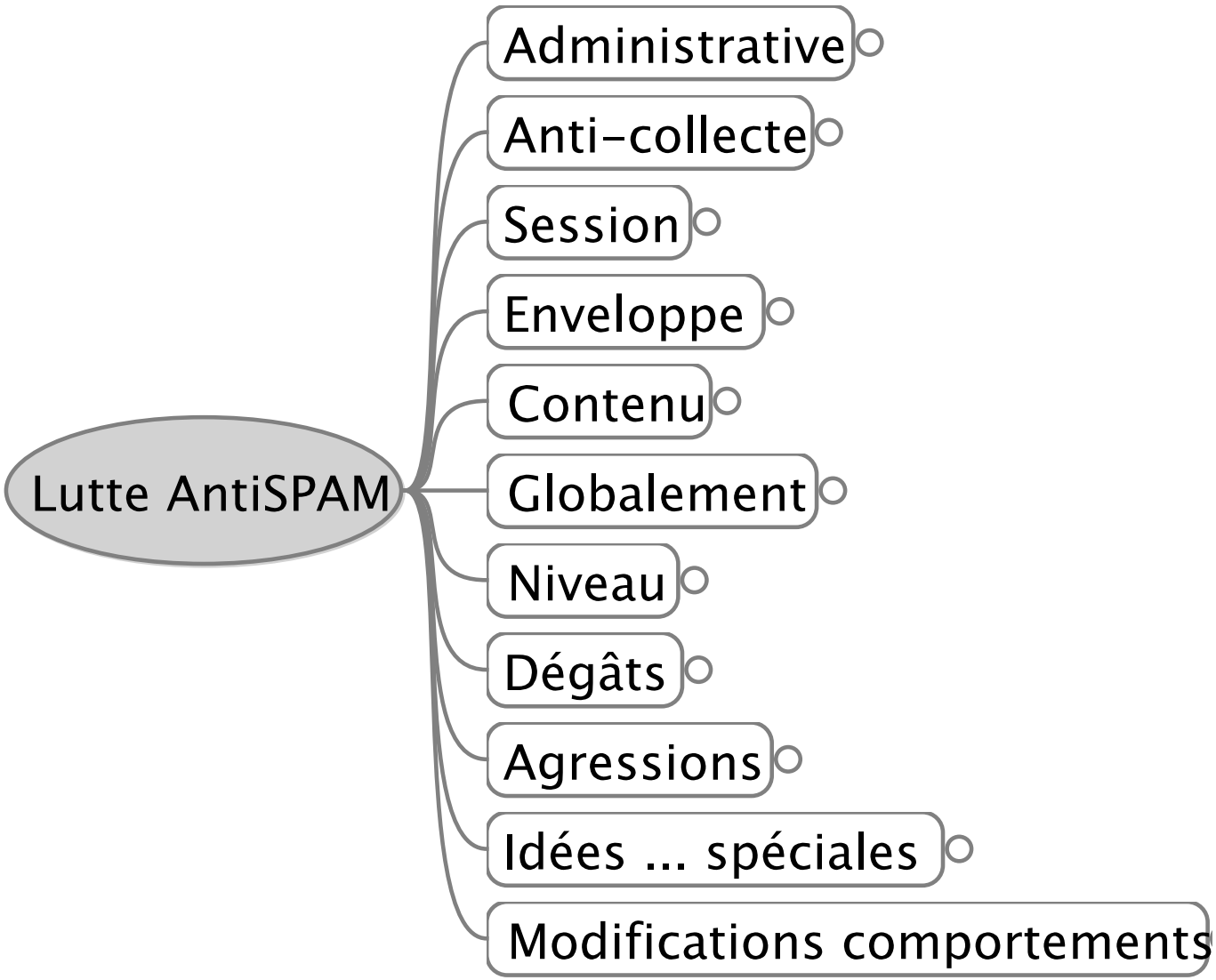
Coût d'une campagne de SPAM

- Location de 10000 machines pour une heure : 50 €
- Base de 20 millions d'email : 250 €

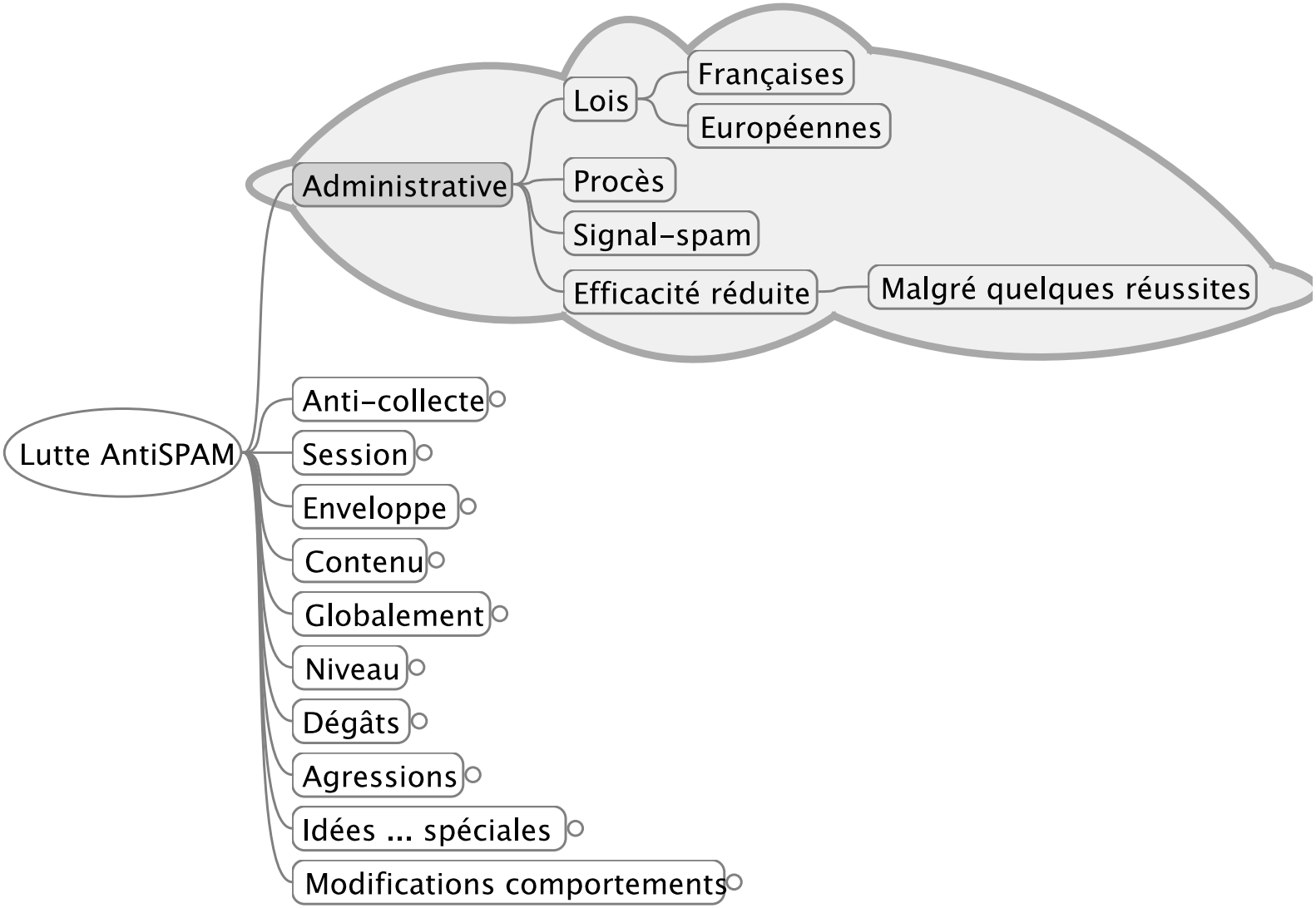
Bénéfice d'une campagne

- Prix d'une montre : 150 €
- Viagra : 100 €les 30 pilules
- Taux de retour estimé : 0.1%
- Campagne "normale" : 1 million de mails

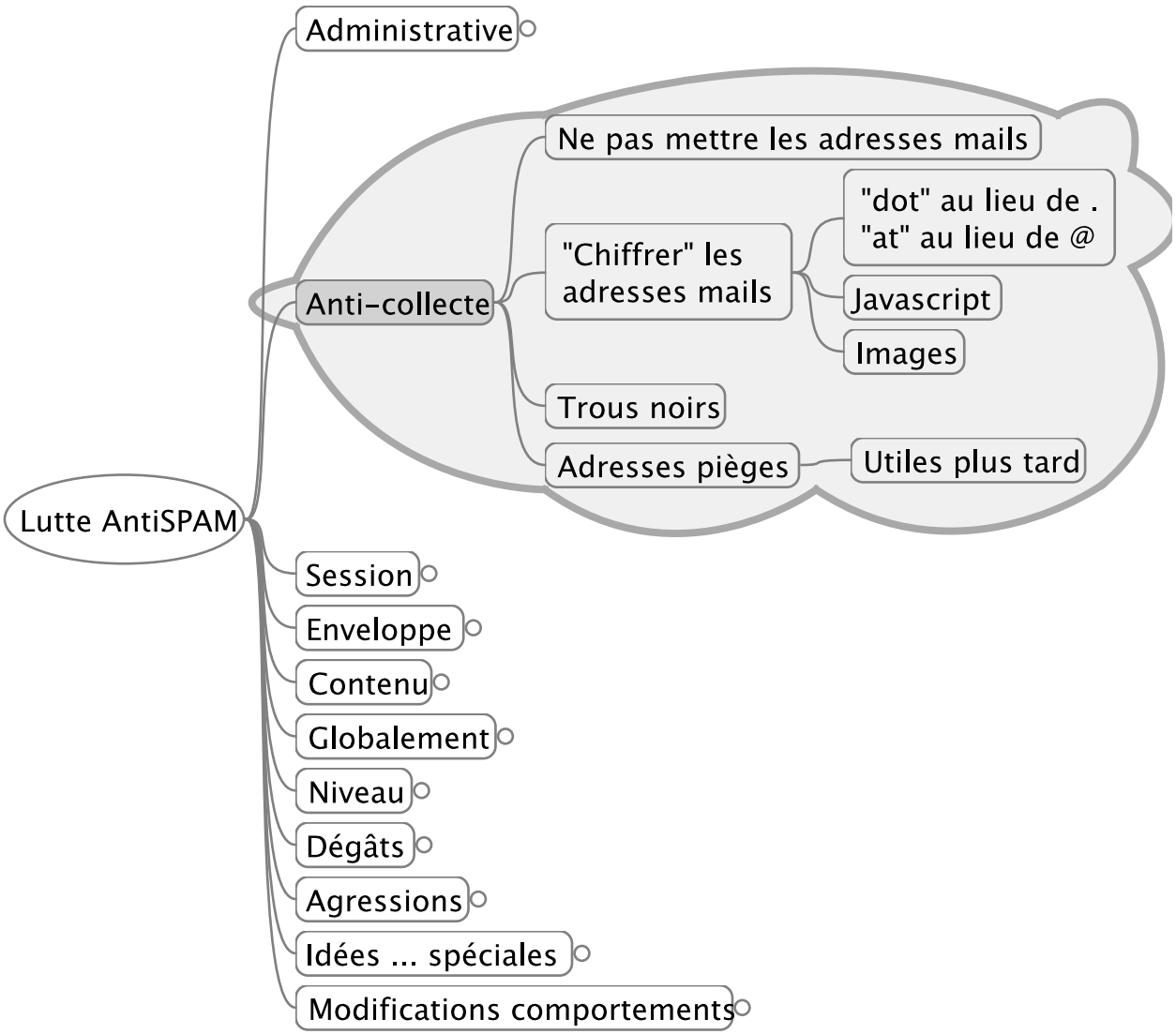
SPAM : où en est la lutte ?



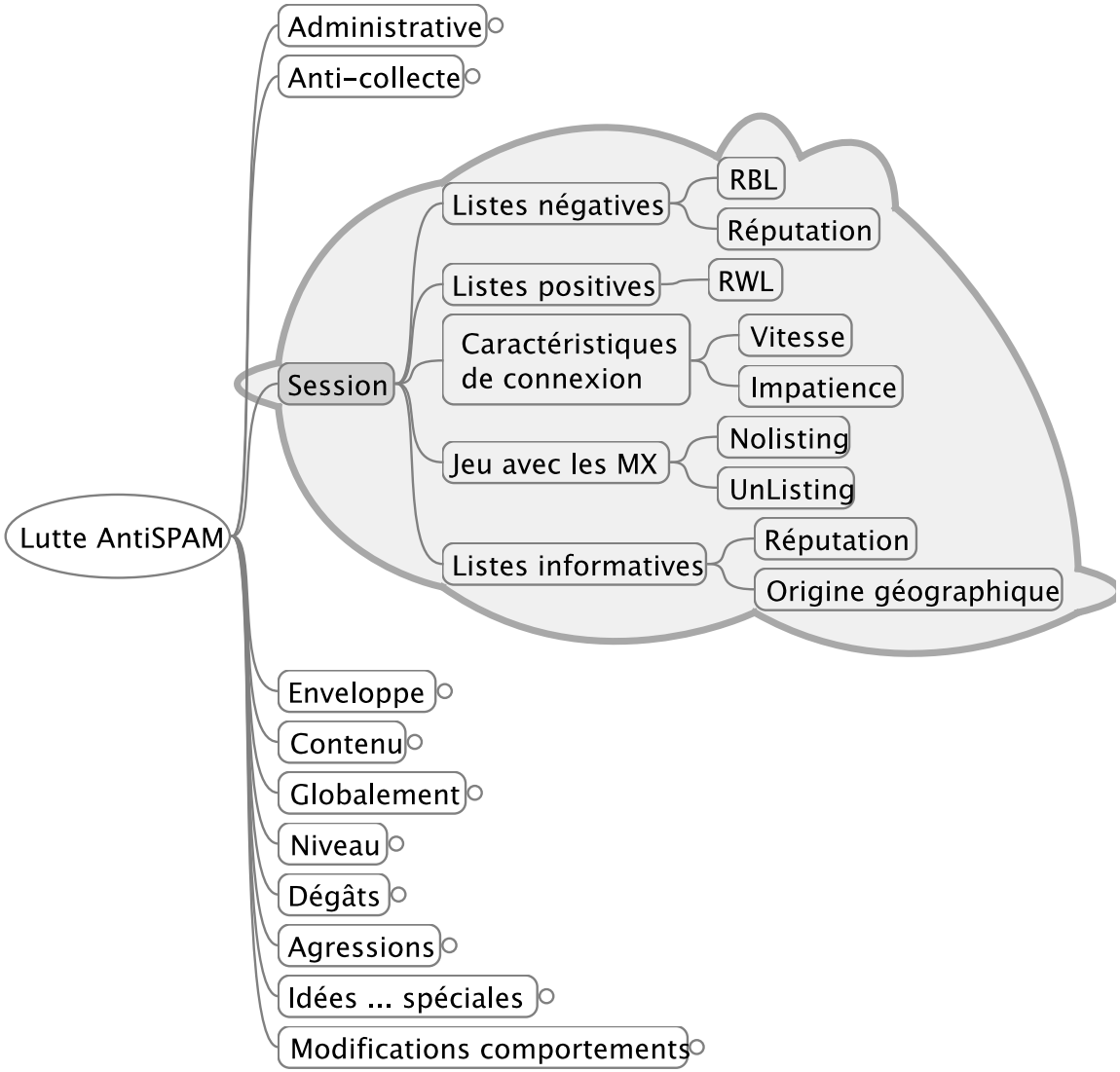
SPAM : Lutte administrative



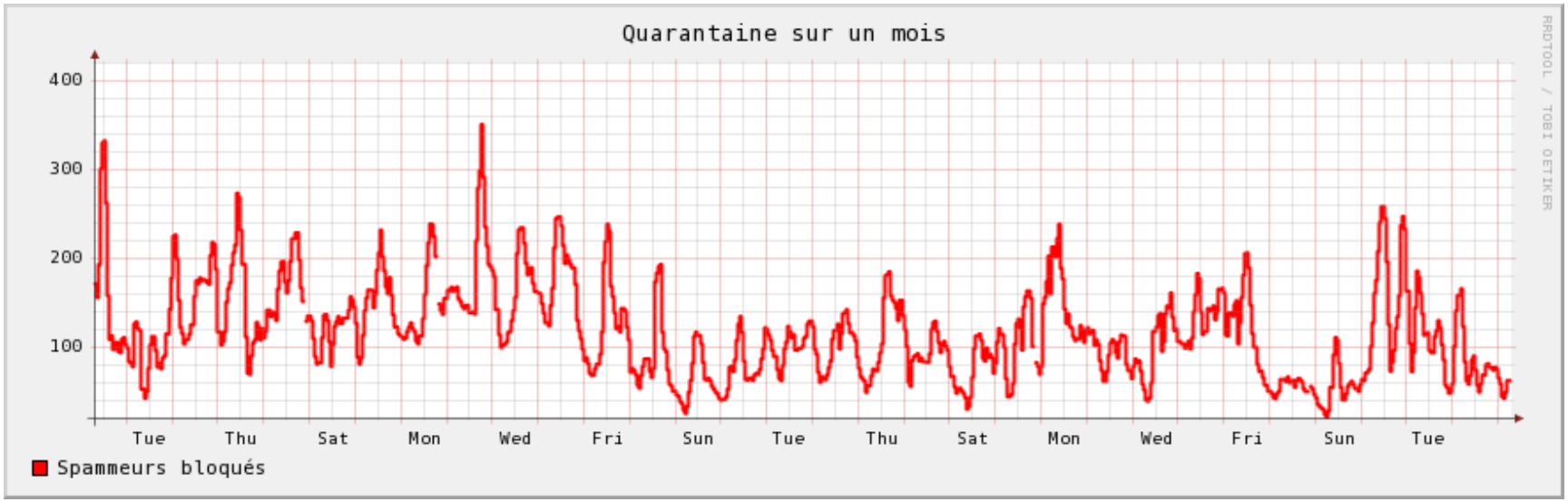
SPAM : Lutte anti collecte



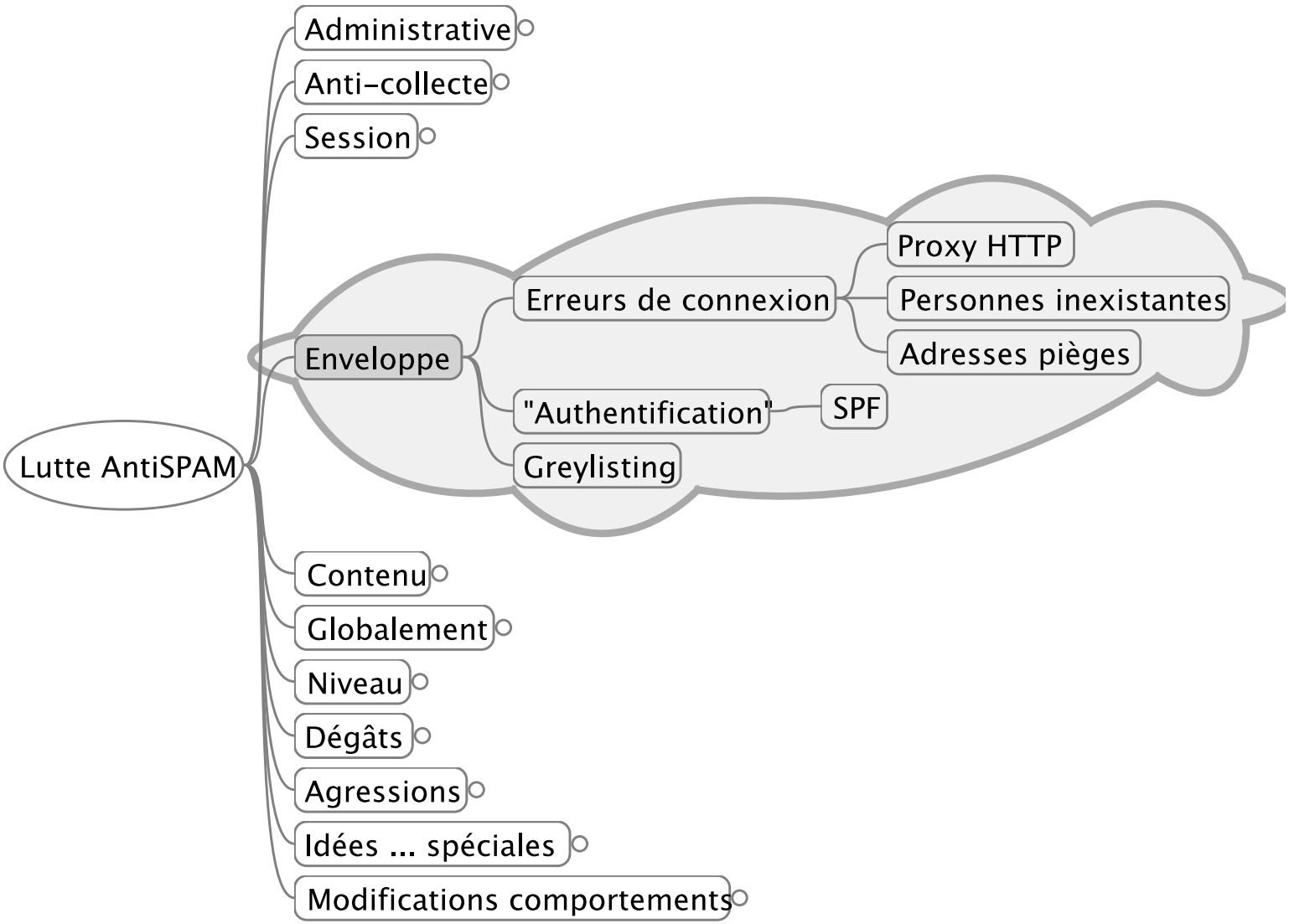
SPAM : Lutte au niveau de la session



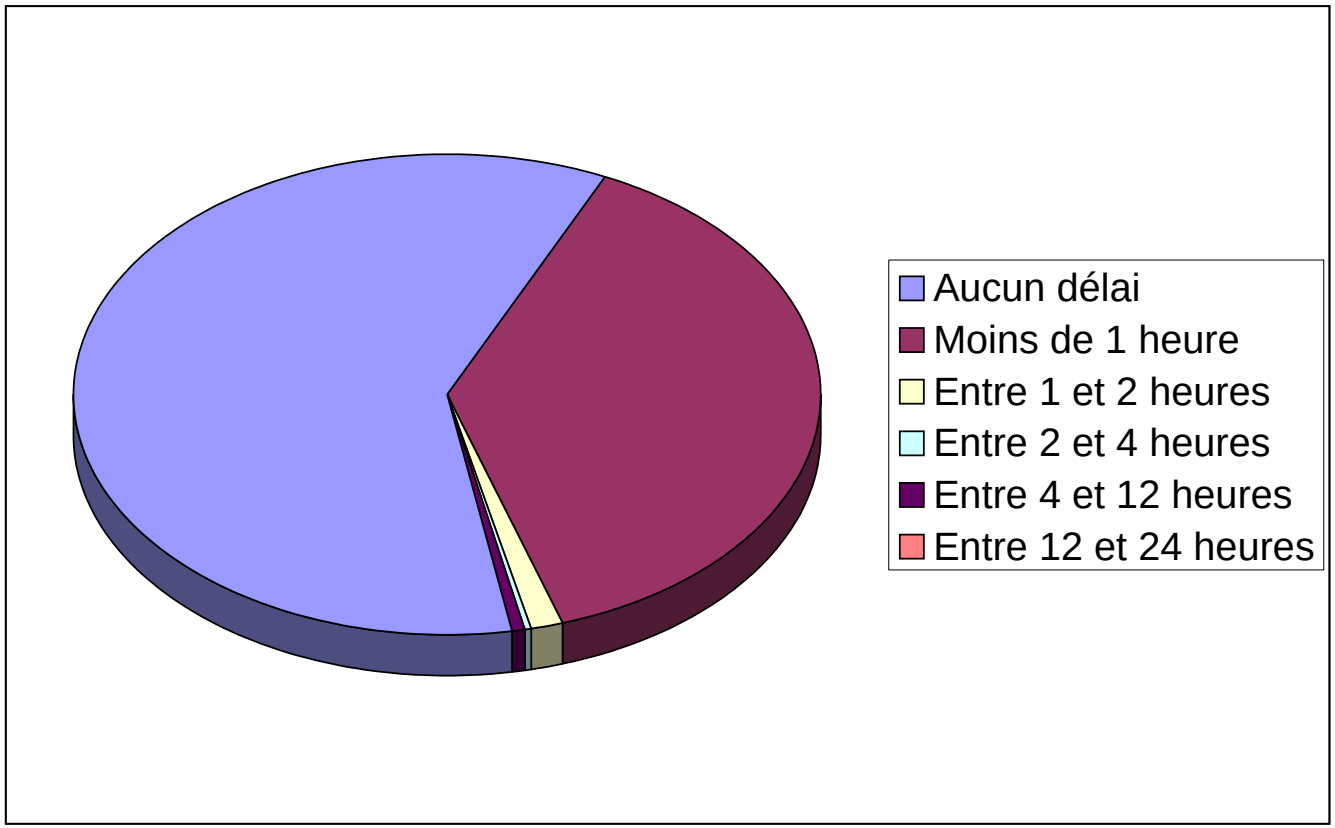
SPAM : Lutte au niveau de la session



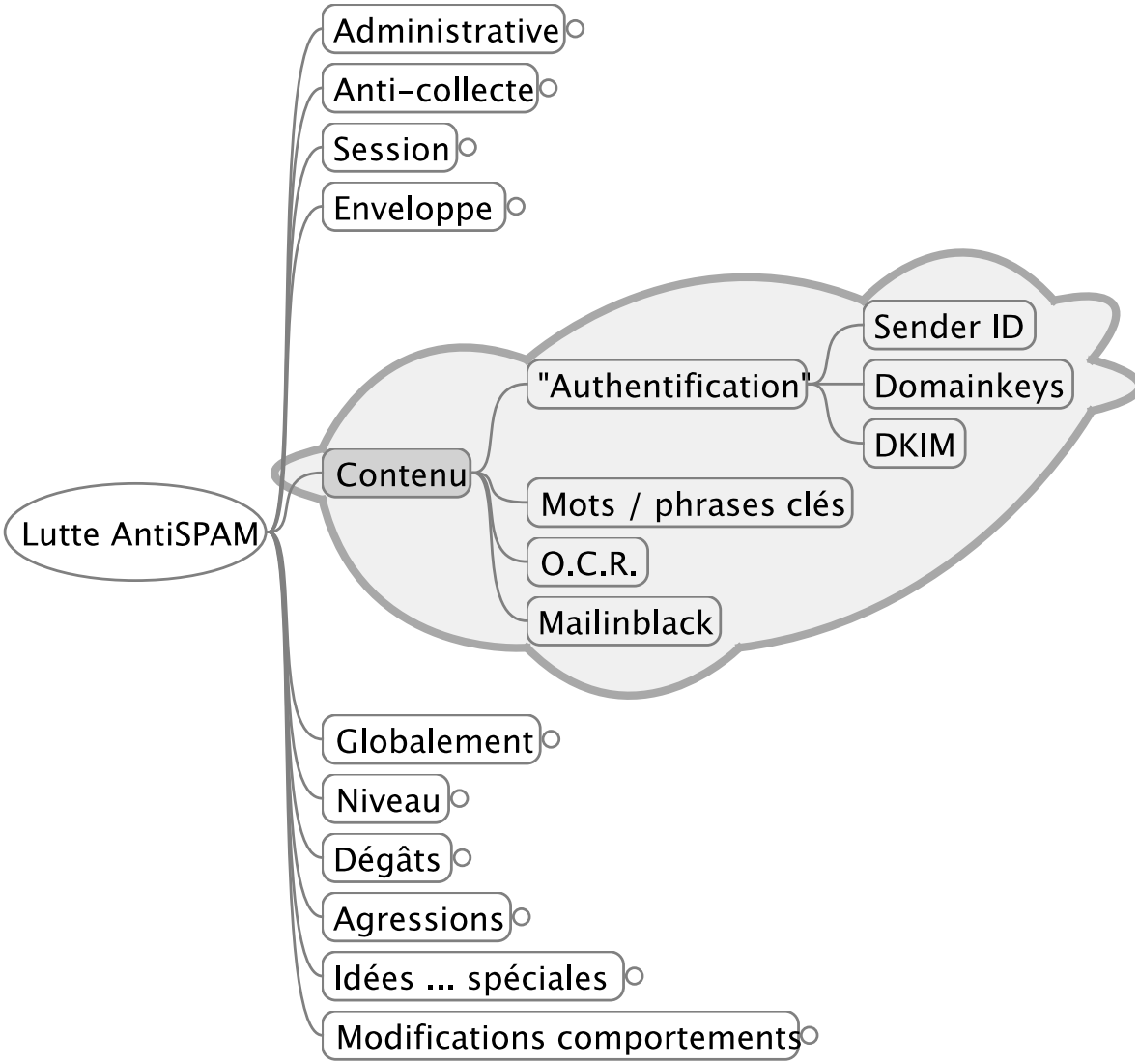
SPAM : Lutte au niveau de l'enveloppe



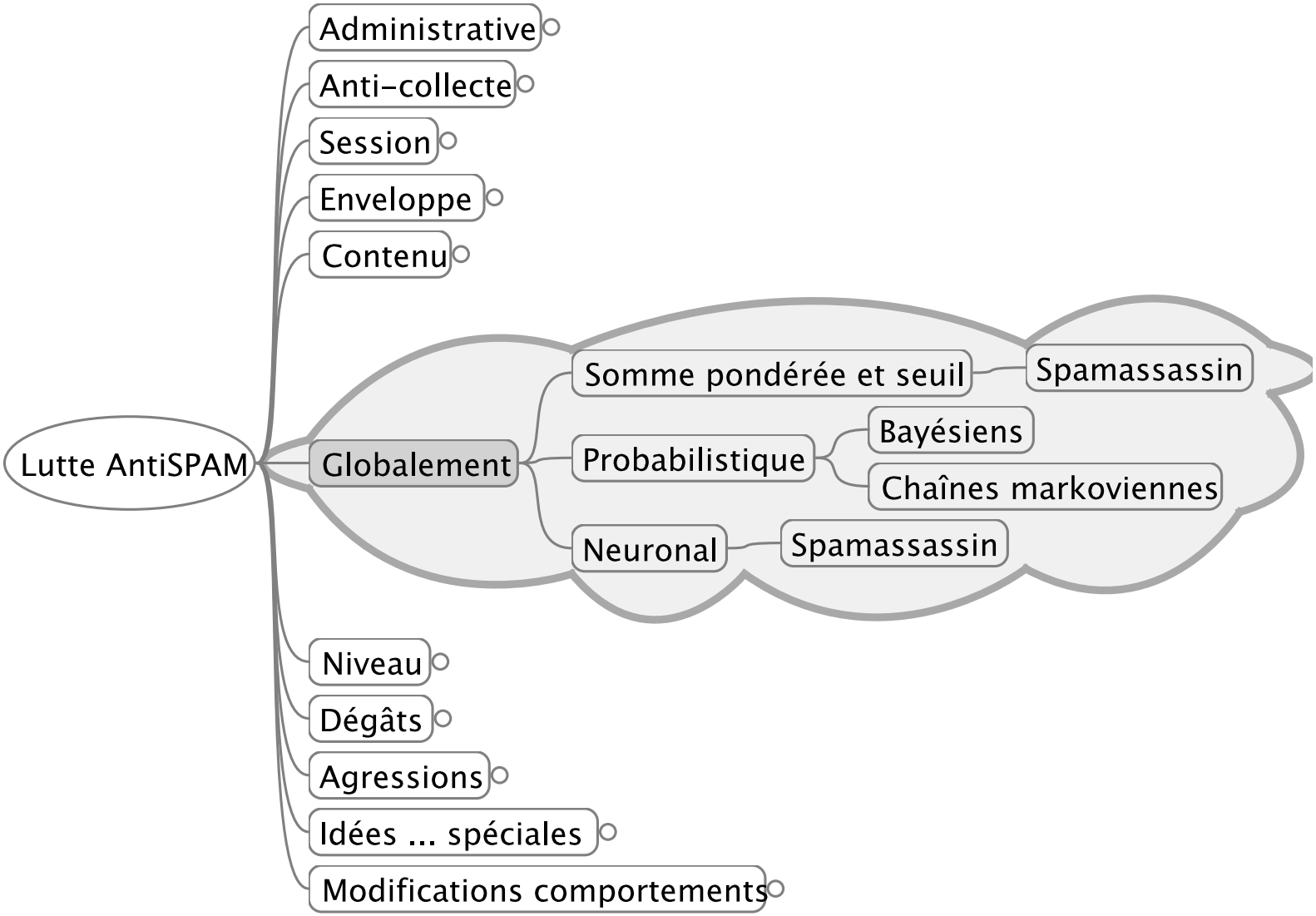
SPAM : retard du greylisting



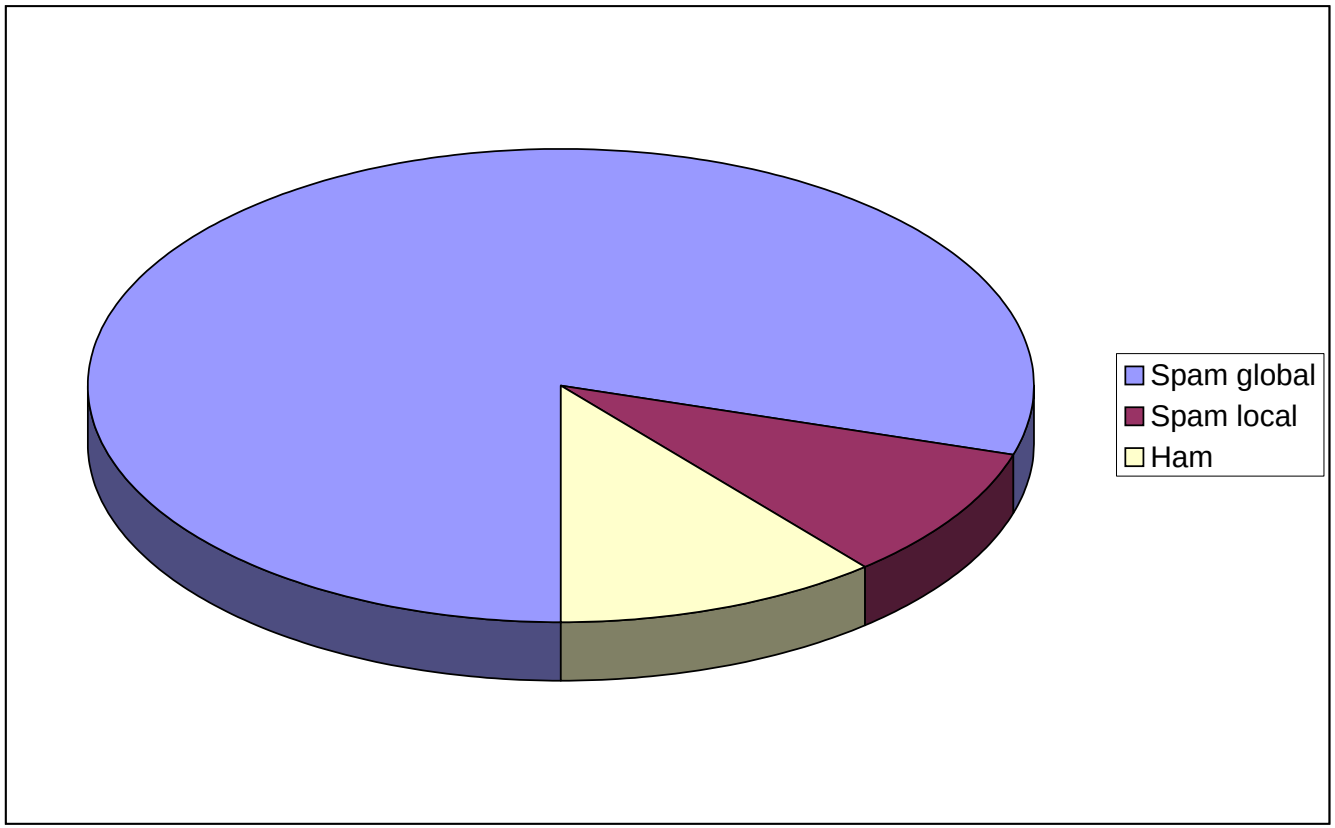
SPAM : Lutte au niveau du contenu



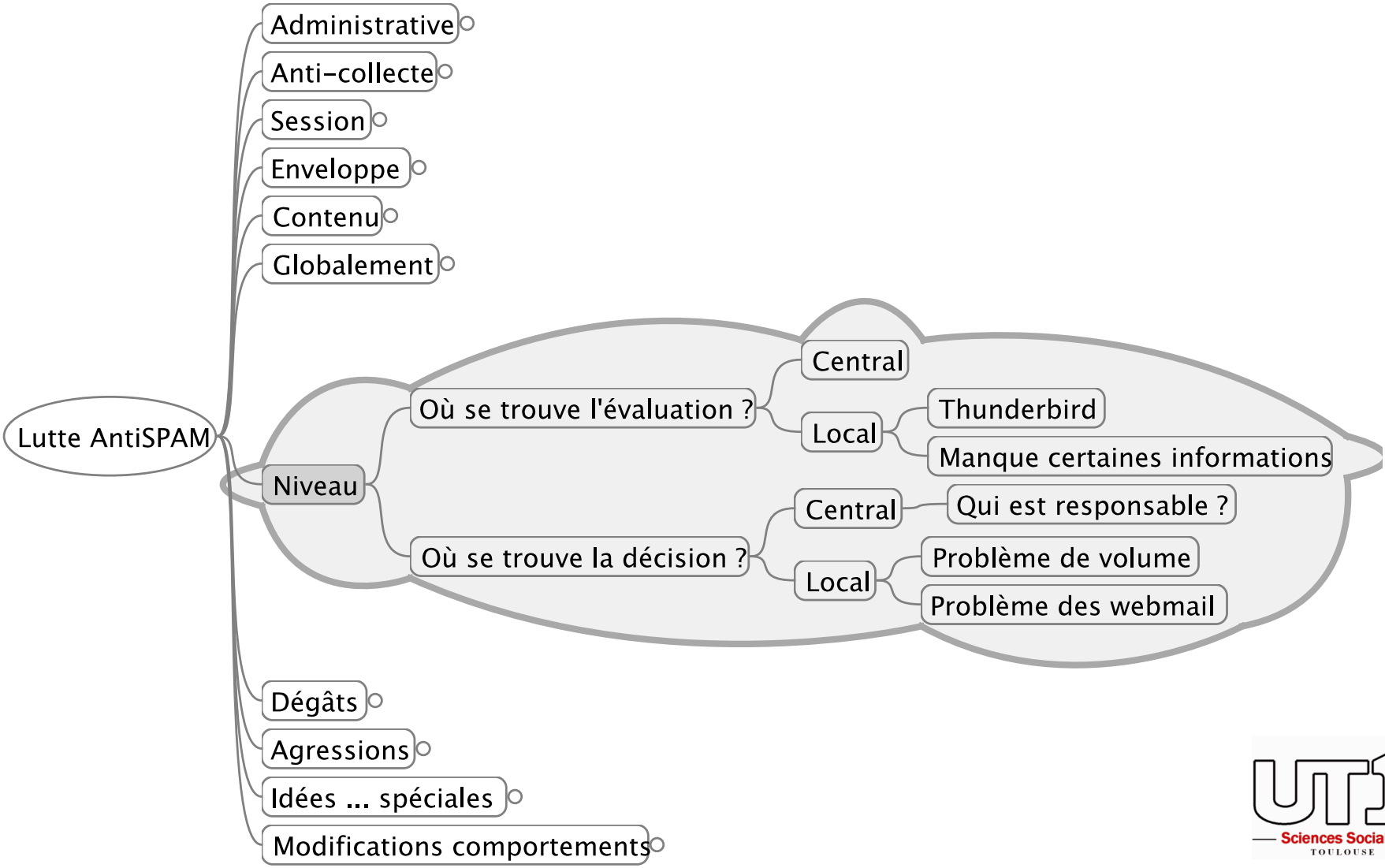
SPAM : Coordination des critères



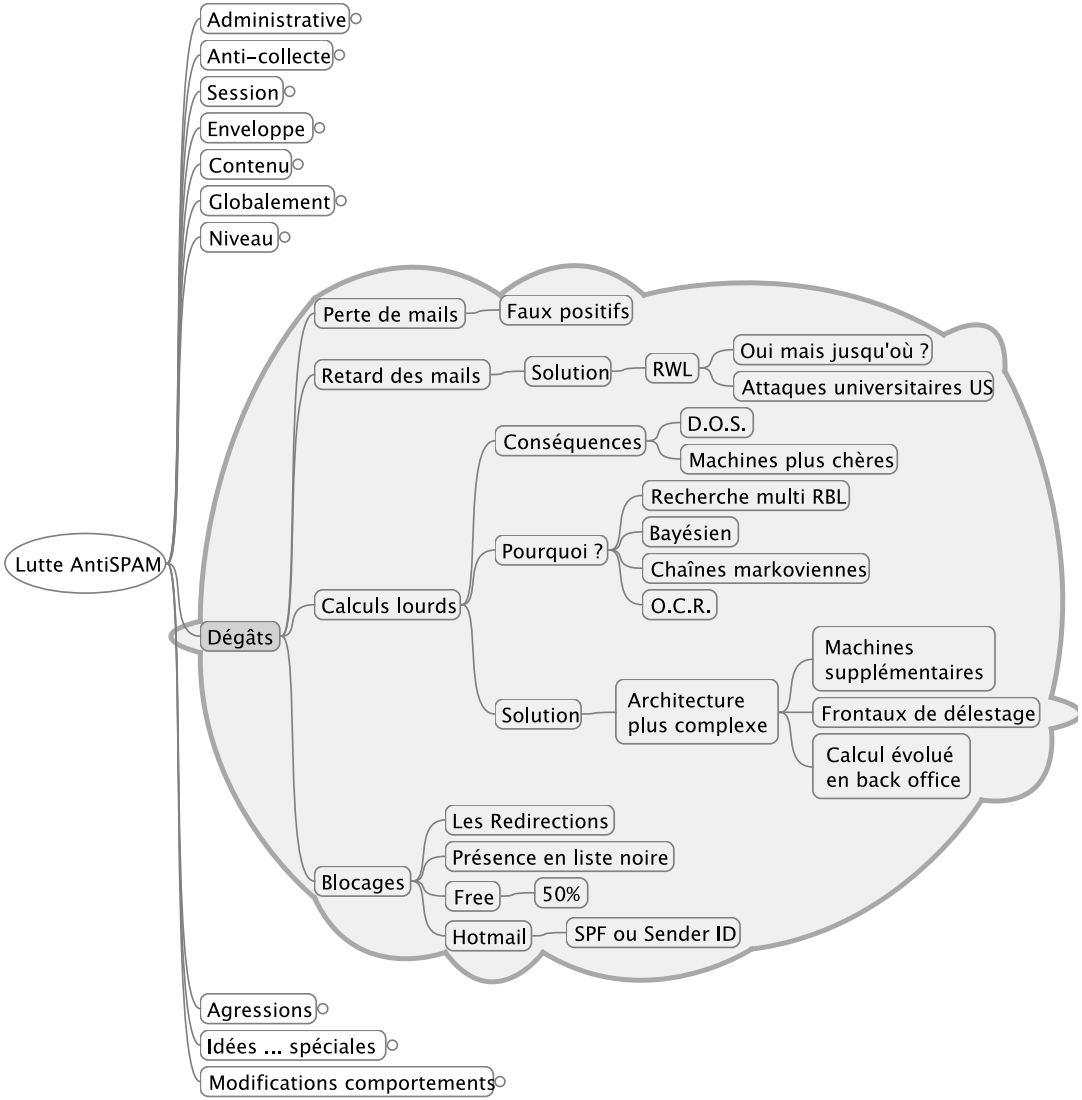
SPAM : Répartition des blocages



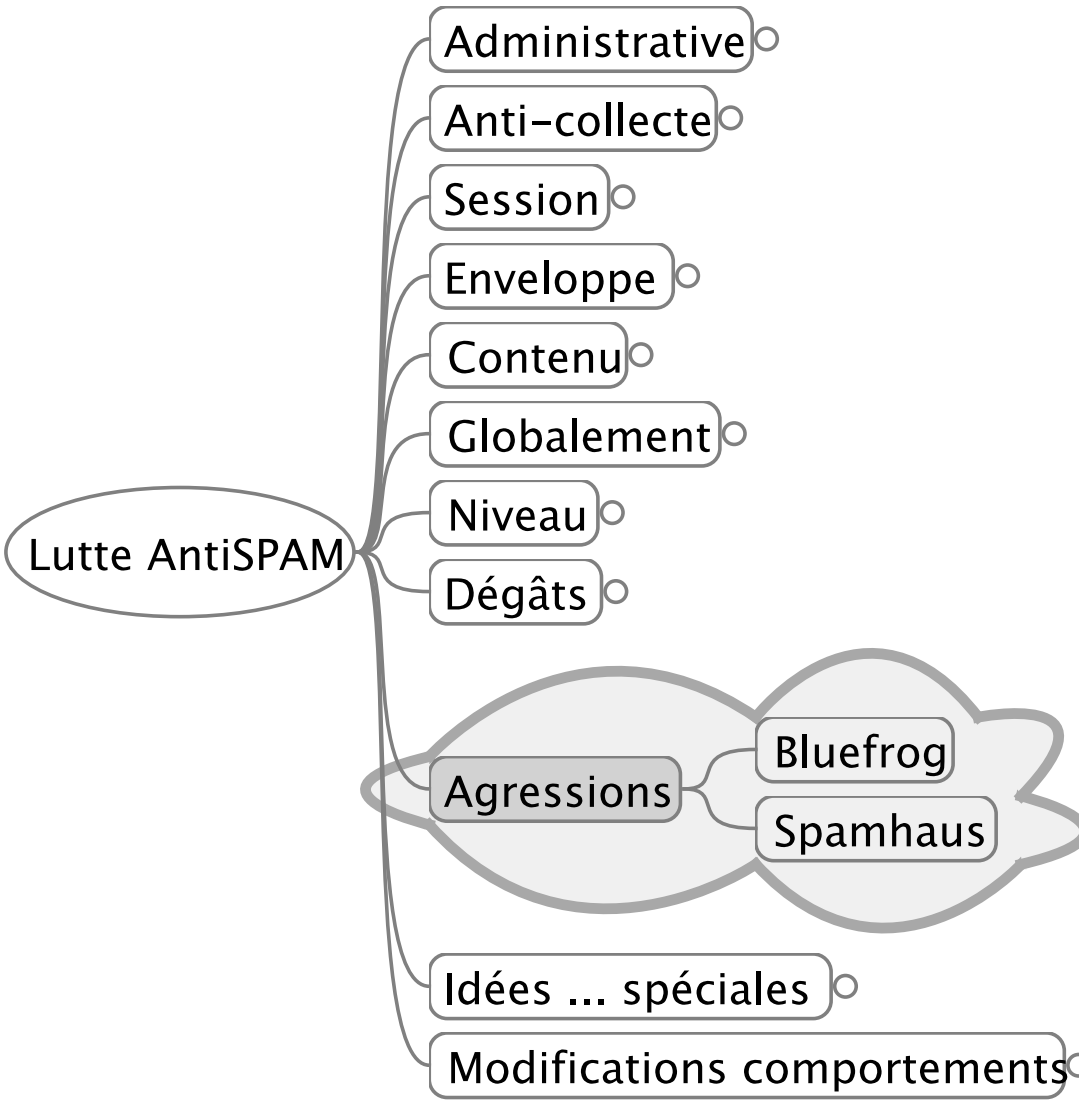
SPAM : Niveau de la lutte



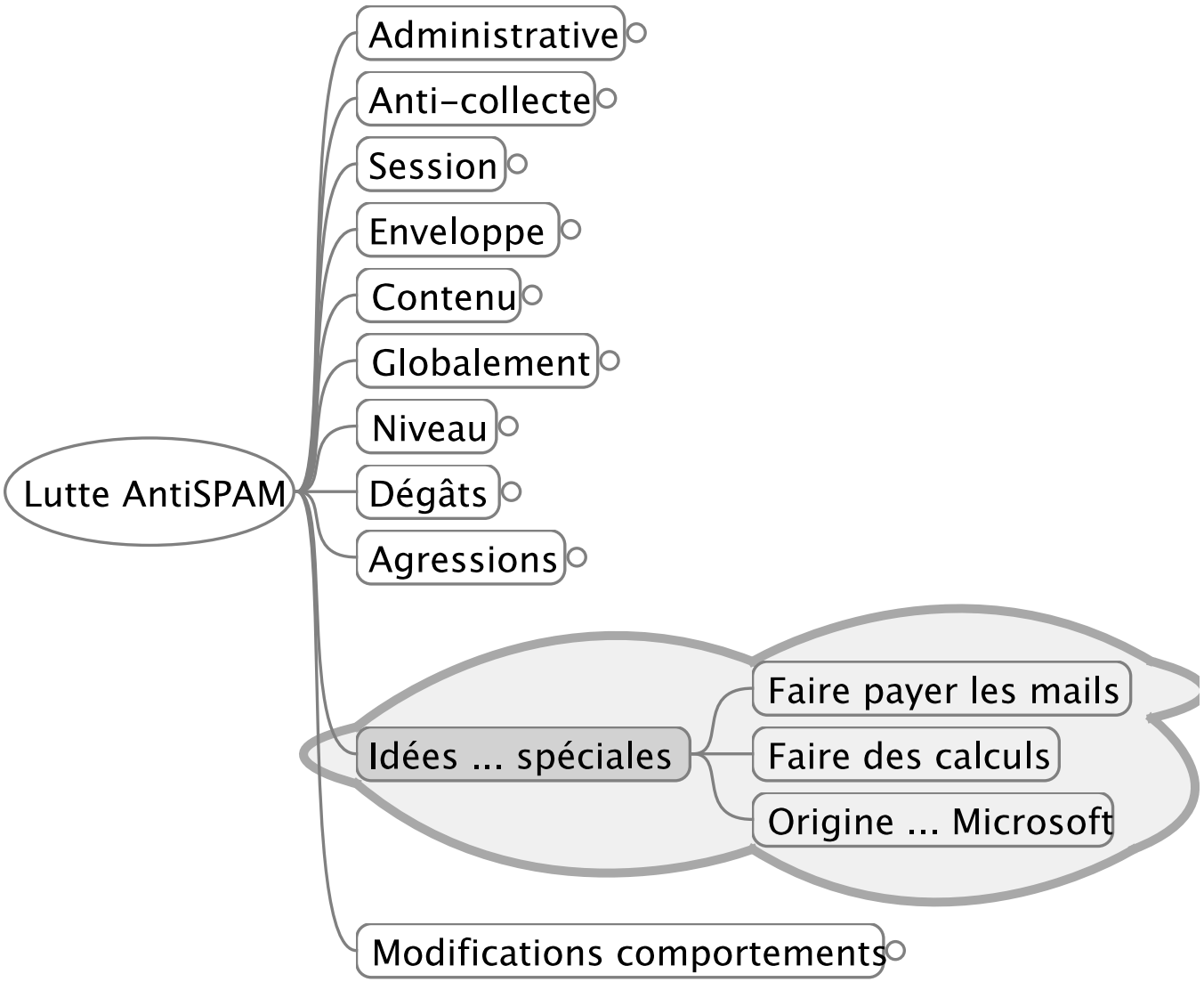
SPAM : Les dégâts de la lutte antispam



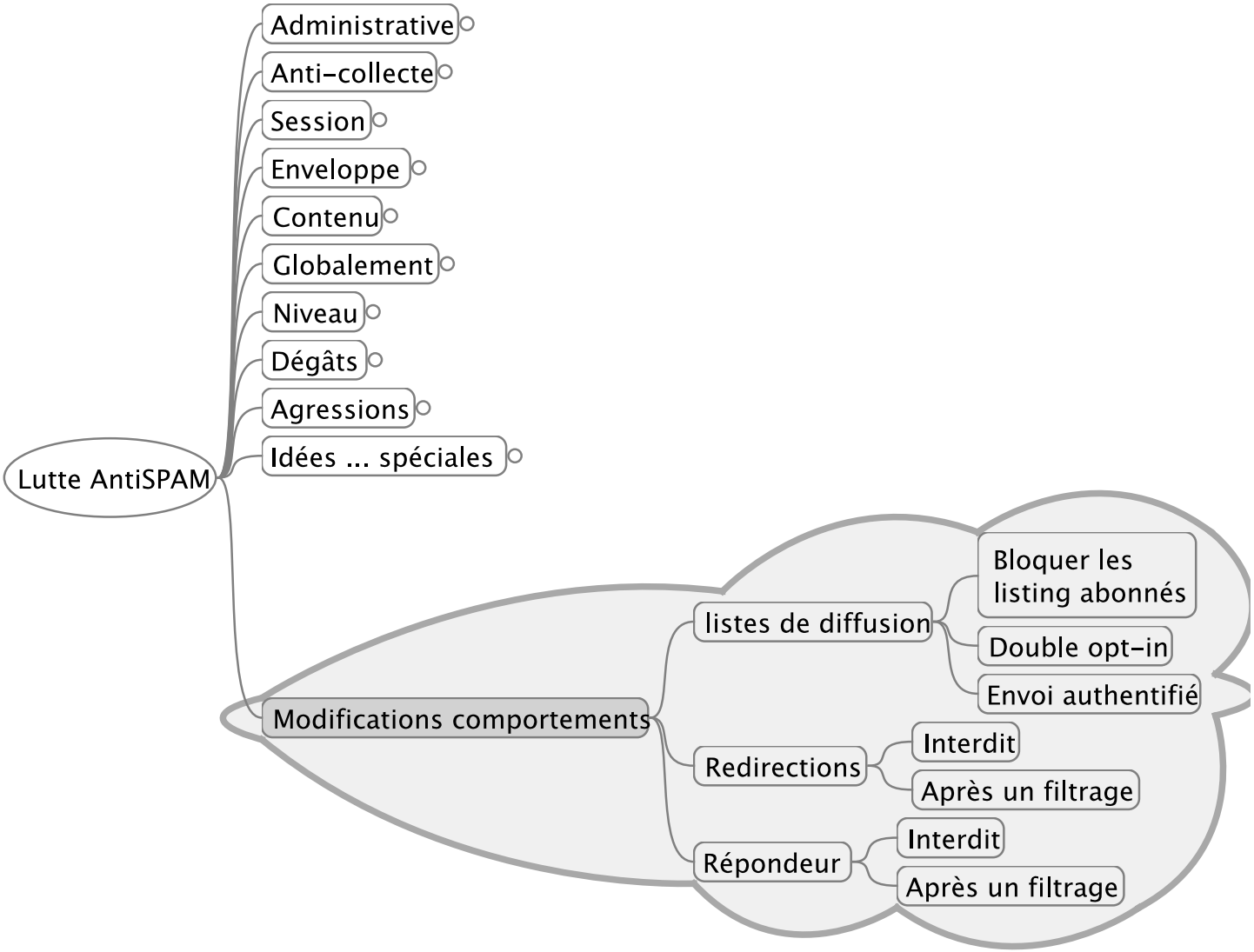
SPAM : Les agressions



SPAM : Quelques idées ... spéciales



SPAM : Les modifications de comportements



Plus de monde touché

Du fait du contournement des protections "initiales", plus de personnes sont touchées.

- Moins habituées,
- Donc moins compétentes,
- Avec de mauvais réflexes (désabonnement, clic sur les virus, etc.),
- Plus disparates (difficulté pour les solutions "globales"),
- Et dont les particularités ne sont pas prises en compte par les logiciels (vocabulaire métier),
- Et acceptant moins les contraintes.

Modifier l'approche

- Diminuer la compétence nécessaire,
- Individualiser la détection avec des dictionnaires spécifiques,
- Individualiser la décision,
- Améliorer la détection, en ne travaillant plus sur les "mots".

DSPAM

DSPAM

- disponible sur http ://dspam.nuclearelephant.com
- un outil libre
- écrit en C
- dernière version : 3.8.0
- qui étend les capacités traditionnelles des bayésiens

Il est doté de nombreuses particularités qui ont pour but d'améliorer le filtrage des spams, mais surtout d'en faciliter l'accès.

DSPAM : les chaines markoviennes

DSPAM fonctionne, par défaut, par le principe des chaines markoviennes.
En gros, c'est un bayésien qui ne travaille pas avec des mots, mais des suites de mots, par exemple :

Exemples

- X-Test-UT1*RBL+xbl.spamhaus.org
- been+approved
- Received*for+<abuse

Les dictionnaires individuels

DSPAM fonctionne avec des dictionnaires individuels. Ceci va avoir plusieurs conséquences :

- Meilleure adaptation au profil de l'utilisateur,
- Un apprentissage individuel,
- La possibilité de partager entre utilisateurs les dictionnaires
  - Sous forme d'un groupe de partage
  - Sous forme d'un dictionnaire "initial" (tant que le dictionnaire individuel est insuffisant)

Les actions possibles

DSPAM peut faire plusieurs actions quand il détecte un SPAM. Chaque action est individualisée.

- Ajouter l'en-tête "X-DSPAM-Result" avec la valeur DSPAM ou innocent.
- Ajouter une information dans le sujet.
- Mettre en quarantaine sur le serveur (hors procédé type procmail, ou sieve)

L'interface web

DSPAM fournit une interface web très complète pour

- Modifier le comportement de DSPAM (le paramétrage)
- Corriger les erreurs
- Afficher les performances (faux positifs et faux négatifs)
- Gérer la quarantaine
- Visualiser des graphes

Une version "administrateur" existe.

Web : Performances

Statistical SPAM Protection for prigent@univ-tlse1. Change

Performance | Préférences | Alertes | Quarantaine (Empty) | Analyse | Historique | Administrative Suite

- Si vous recevez un message dans votre logiciel de messagerie qui n'a pas été détecté par le filtre, veuillez le renvoyer à l'adresse **mail.incorrect@univ-tlse1.fr** afin qu'il soit analysé et repéré comme SPAM.
- Si vous recevez un message dans votre logiciel de messagerie qui a été incorrectement détecté par le filtre comme SPAM, veuillez le renvoyer à l'adresse **mail.correct@univ-tlse1.fr** afin qu'il soit analysé et repéré comme HAM.

Ceci améliorera graduellement la précision du filtre.

Statistiques de performance - Mon Jan 29 7:21:28 2007

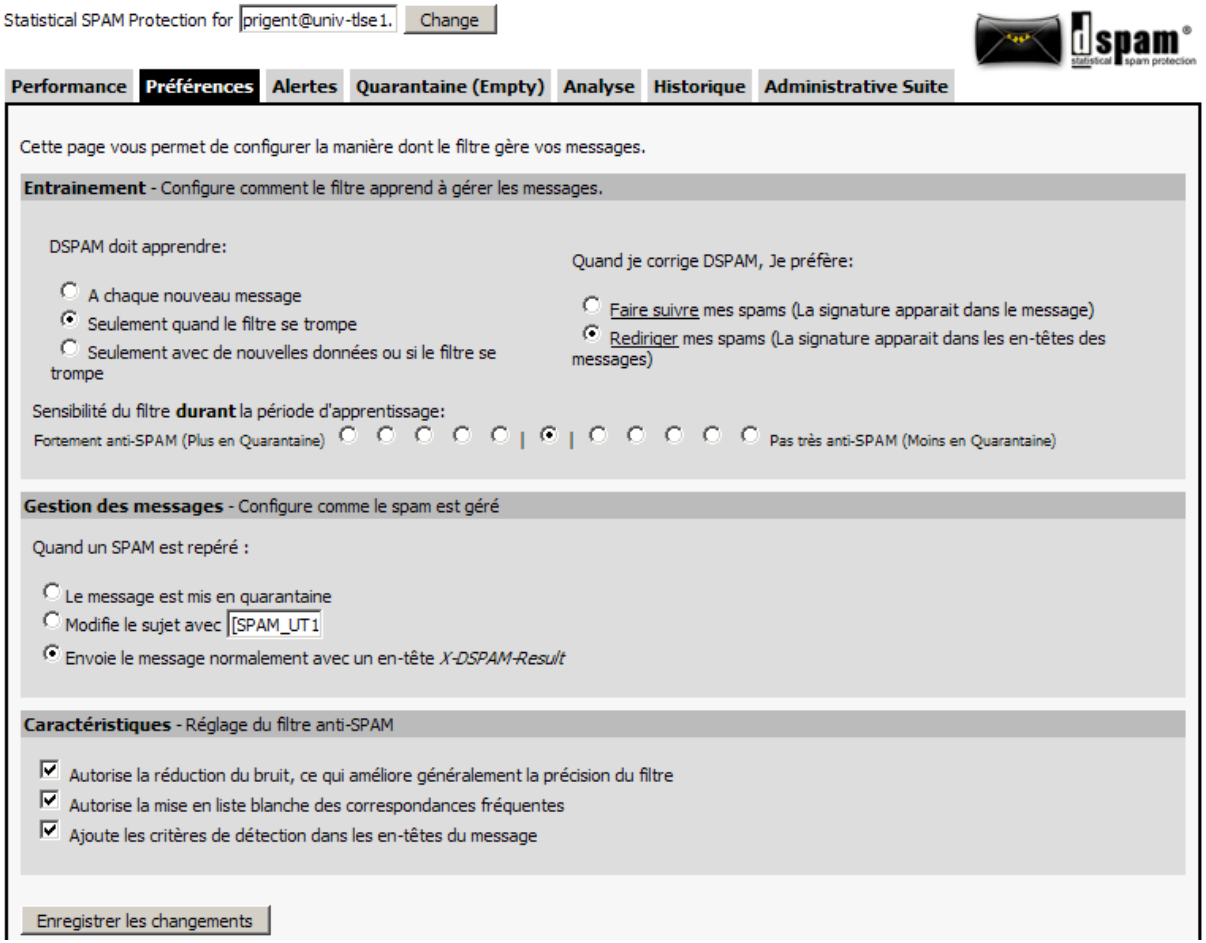
Métrique		Formule de calcul
Efficacité globale (depuis la dernière RAZ)	**99.889%**	(SPAM détectés + messages corrects délivrés) / Nombre total de messages
Identification de SPAM (depuis la dernière RAZ)	**99.887%**	(Taux de détection de SPAM uniquement)
Spam ratio (sur le nombre total de message)	**30.916%**	Total de SPAM (Détectés ou pas) / Nombre total de messages

	SPAM	HAMS
Depuis la dernière remise à zéro	51 ratés	117 ratés
	44893 détectés	106206 délivrés
	99.887% détectés	0.110% ratés
Total traités par le filtre	142 ratés	304 ratés
	57150 détectés	127719 délivrés
Depuis un corpus	50 fed	374 fed

Remise à zéro | Tweak -1

DSPAM is a registered trademark of Deep Logic, Inc.. Copyright © 2005, Deep Logic, Inc.

Web : Préférences



Web : Quarantaine

Statistical SPAM Protection for prigent@univ-tlse1.fr Change

dspam statistical spam protection

Performance | Preferences | Alertes | Quarantaine (15) | Analyse | Historique | Administrative Suite

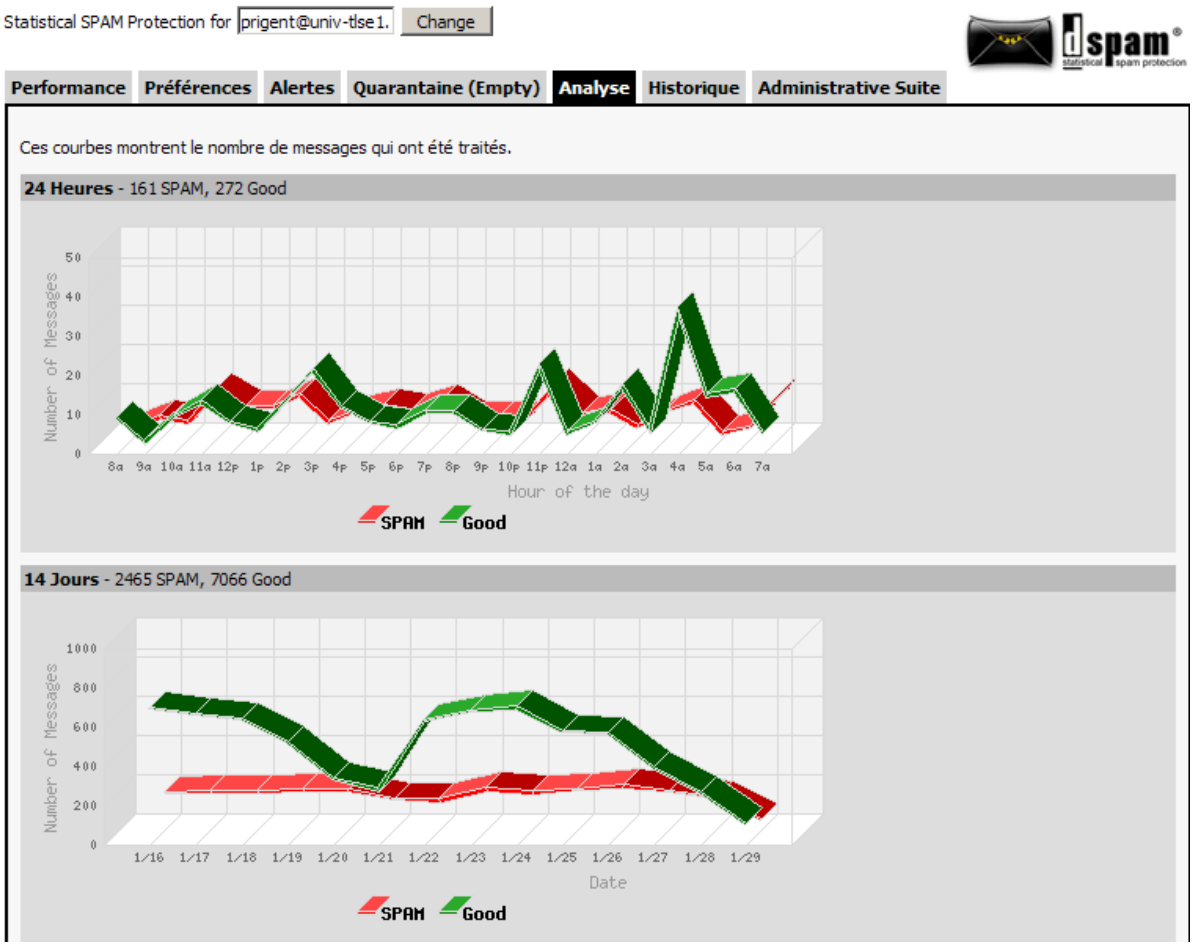
Les messages ci dessous n'ont pas été reçus par votre logiciel de courrier électronique car ils ont été repérés comme spam. Cliquez sur la ligne Sujet pour voir le message ou choisissez une option de tri pour modifier l'ordre d'affichage. Utilisez les boites à cocher et **Envoyer** pour libérer les messages que vous souhaitez lire ou utilisez **Effacer Tout** pour vider la quarantaine.

Envoyer | Effacer | Effacer Tout | Sort by: **Rating** | Date | Subject | From

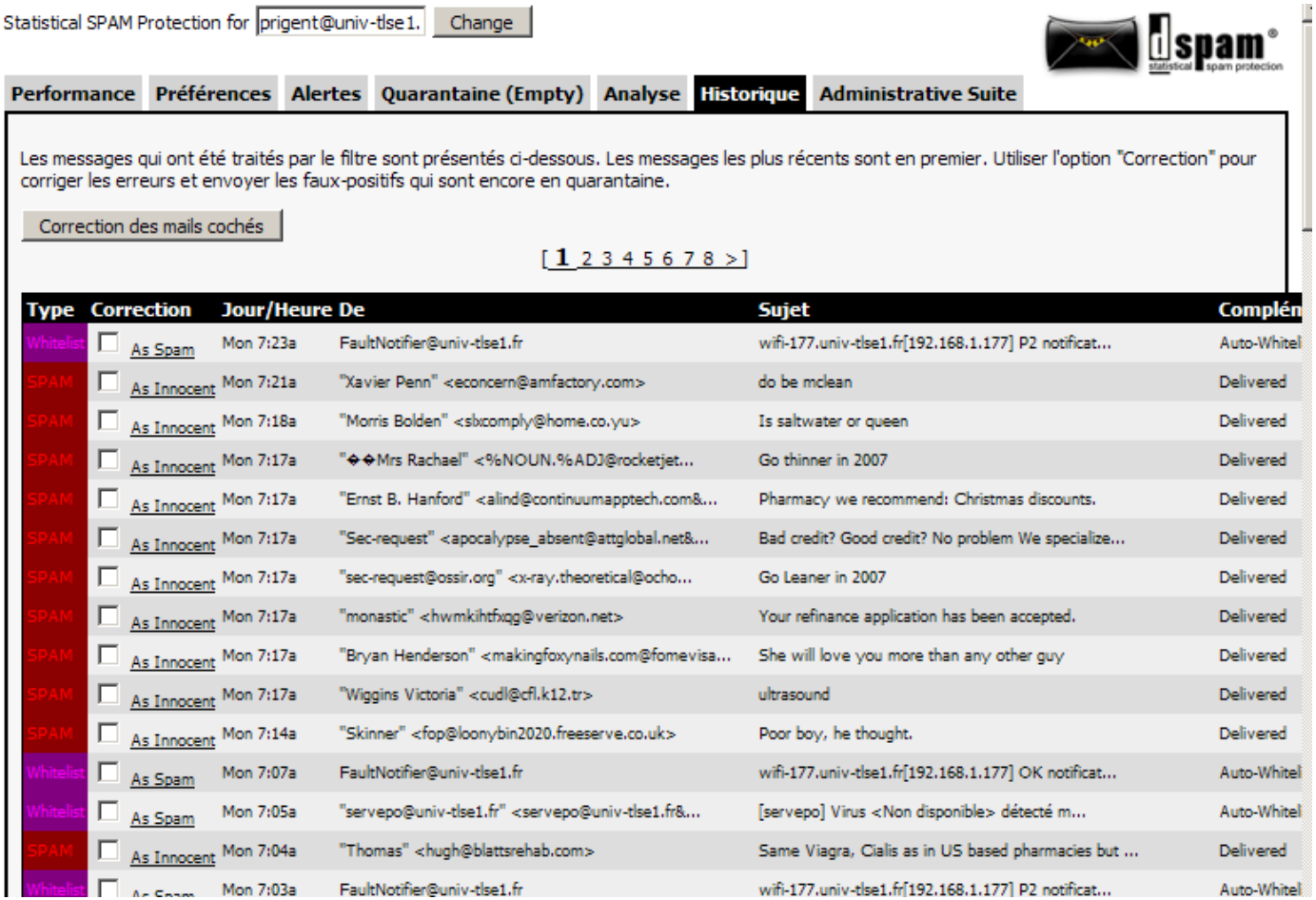
Note	Date	Expediteur	Sujet
70%	Jan 29 07:46a	Millie A. Clifton <zic@glmusicals.com>	terms
76%	Jan 29 07:44a	Janet C. Martinez <gby@joomlaserbia.com>	revere
85%	Jan 29 09:17a	"Tivadar Kershaw" <brannumu@jasanco.com>	Re: pigRX
99%	Jan 29 08:17a	"anish mohamad" <anishmohamad@yahoo.co.in>	Obtain the career you have always wanted with the ...
99%	Jan 29 08:30a	Bank of America <noreply@bankofamerica.com>	Security: Your Online Banking Account is Blocked
99%	Jan 29 08:44a	"Marina Farr" <lnwaisjdhcck@snail-mail.net>	best prices for impotence drug$!
99%	Jan 29 07:49a	Central Bank <eebi_cbn@mac.com>	YOUR OVER DUE INHERITANCE FUND
99%	Jan 29 07:47a	"Rene Meiers" <ipesrol@options-st.com>	backlog vacuum
99%	Jan 29 08:17a	"Mark Christie" <SMSHDNCTDAXH@epomail.com>	SPUR-M
100%	Jan 29 08:04a	"Gamble Edmund" <ombp@malkol.cz>	CropsChoosing which crops to grow depends on many ...
100%	Jan 29 08:17a	"Cyrus G. Rivera" <qui@jsmtel.com>	Do not take a report over the telephone as being t...
100%	Jan 29 07:59a	"cocacola_rep@terra.es" <cocacola_rep@terra.es>	THE COCA COLA COMPANY SEASON PRIZE NOTIFICATION
100%	Jan 29 08:06a	"cocacola_rep@terra.es" <cocacola_rep@terra.es>	THE COCA COLA COMPANY SEASON PRIZE NOTIFICATION

UT1 Sciences Sociales TOULOUSE

Web : Analyse


Statistical SPAM Protection for prigent@univ-tlse1.
Change
dspam
Performance
Préférences
Alertes
Quarantaine (Empty)
Analyse
Historique
Administrative Suite
Ces courbes montrent le nombre de messages qui ont été traités.
24 Heures - 161 SPAM, 272 Good
Number of Messages
Hour of the day
SPAM
Good
14 Jours - 2465 SPAM, 7066 Good
Number of Messages
Date
SPAM
Good

Web : Historique



Statistical SPAM Protection for prigent@univ-tlse1. Change

dspam statistical spam protection

Performance | Préférences | Alertes | Quarantaine (Empty) | Analyse | Historique | Administrative Suite

Les messages qui ont été traités par le filtre sont présentés ci-dessous. Les messages les plus récents sont en premier. Utiliser l'option "Correction" pour corriger les erreurs et envoyer les faux-positifs qui sont encore en quarantaine.

Correction des mails cochés

[1 2 3 4 5 6 7 8 >]

Type	Correction	Jour/Heure	De	Sujet	Complén
Whitelist	As Spam	Mon 7:23a	FaultNotifier@univ-tlse1.fr	wifi-177.univ-tlse1.fr[192.168.1.177] P2 notificat...	Auto-Whitel
SPAM	As Innocent	Mon 7:21a	"Xavier Penn" <econcern@amfactory.com>	do be mclean	Delivered
SPAM	As Innocent	Mon 7:18a	"Morris Bolden" <sbxcomply@home.co.yu>	Is saltwater or queen	Delivered
SPAM	As Innocent	Mon 7:17a	"◆◆Mrs Rachael" <%NOUN.%ADJ@rocketjet...	Go thinner in 2007	Delivered
SPAM	As Innocent	Mon 7:17a	"Ernst B. Hanford" <alind@continuumapptech.com&...	Pharmacy we recommend: Christmas discounts.	Delivered
SPAM	As Innocent	Mon 7:17a	"Sec-request" <apocalypse_absent@attglobal.net&...	Bad credit? Good credit? No problem We specialize...	Delivered
SPAM	As Innocent	Mon 7:17a	"sec-request@ossir.org" <x-ray.theoretical@ocho...	Go Leaner in 2007	Delivered
SPAM	As Innocent	Mon 7:17a	"monastic" <hwmkihtfxqg@verizon.net>	Your refinance application has been accepted.	Delivered
SPAM	As Innocent	Mon 7:17a	"Bryan Henderson" <makingfoxynails.com@fomevisa...	She will love you more than any other guy	Delivered
SPAM	As Innocent	Mon 7:17a	"Wiggins Victoria" <cudl@cfl.k12.tr>	ultrasound	Delivered
SPAM	As Innocent	Mon 7:14a	"Skinner" <fop@loonybin2020.freeserve.co.uk>	Poor boy, he thought.	Delivered
Whitelist	As Spam	Mon 7:07a	FaultNotifier@univ-tlse1.fr	wifi-177.univ-tlse1.fr[192.168.1.177] OK notificat...	Auto-Whitel
Whitelist	As Spam	Mon 7:05a	"servepo@univ-tlse1.fr" <servepo@univ-tlse1.fr&...	[servepo] Virus <Non disponible> détecté m...	Auto-Whitel
SPAM	As Innocent	Mon 7:04a	"Thomas" <hugh@blattsrehab.com>	Same Viagra, Cialis as in US based pharmacies but ...	Delivered
Whitelist	As Spam	Mon 7:03a	FaultNotifier@univ-tlse1.fr	wifi-177.univ-tlse1.fr[192.168.1.177] P2 notificat...	Auto-Whitel

UT1 Sciences Sociales TOULOUSE

Web : Statistiques administrateur

Statistical SPAM Protection for **Administrator (prigent@univ-tlse1.fr)**

System Status | **User Statistics** | **Administration** | **Control Center**

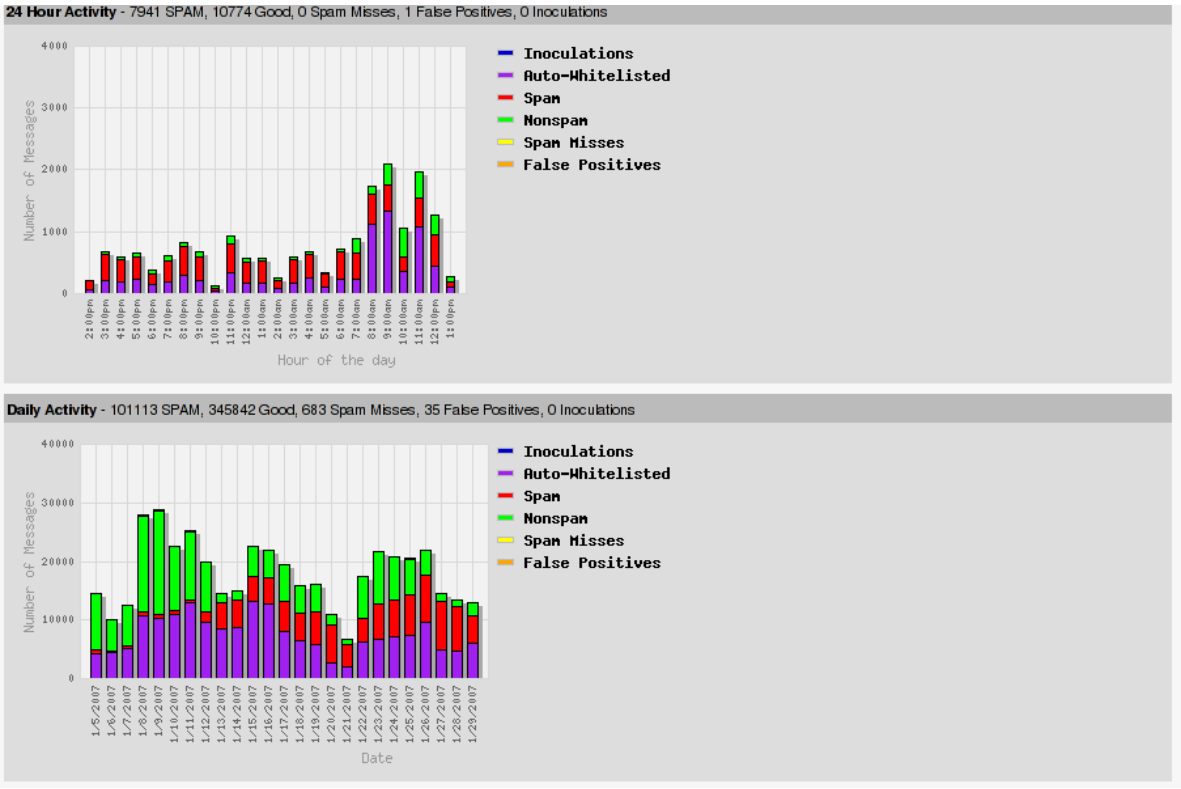
The following graphs and tables summarize the processing done by the filter.

Overview

Messages	Today	This Hour
Spam	4842	82
Good	2258	80
Spam Misses	0	0
False Positives	1	0
Inoculations	0	0
Total	**7101**	**162**

Status	Current Value
Average message processing time	0.239629 sec.
Average throughput	0.20 messages/sec.
DSPAM instances	3 processes
System uptime	13:27:32 up 132 days, 7:14, 3 users, load average: 0.59, 0.30, 0.22
Mail queue length	414 messages

Web : Graphiques administrateur



Ses prérequis

Si DSPAM peut fonctionner avec un grand nombre de procédés de stockage, il ne donne réellement toute sa puissance qu'avec une base de données MySQL.

- Stockage des paramètres personnels (inaccessibles sinon)
- Travail avec des signatures (pour une correction simplifiée)

Sa position

DSPAM peut se placer en tant que relais

- POP, pour avoir des pointes de charge à 8h30 le matin ;-)
- SMTP, avec création d'une boucle MTA / DSPAM / MTA
- LMTP, le plus efficace

Son fonctionnement

Il est identique à un bayésien standard.

Apprentissage

- Avec un pool de mails (2500 mails et au moins 250 spams)
- Récupération de tokens (parfois des phrases)
- Calcul de la spamicité de ces tokens

Utilisation

- Récupération de tokens
- Evaluation de ces tokens
- Réinjection du mail pour réapprentissage
  - Toujours (TEFT)
  - Sur erreur (TOE)
  - Sans entrainement (unlearn)
  - etc.

Son fonctionnement original

Du fait de sa granularité, il va ajouter certains aspects

- Création d'une signature pour chaque message
- Remplissage avec les tokens du message
- Insertion dans le message de la signature
  - Dans les en-têtes si on corrige en redirigeant le message
  - En fin de message si on corrige en le faisant suivre.
- La zone signature reste jusqu'à effacement (procédure externe)
- La correction peut être faite par interface web.

Les en-têtes ajoutés

Spam

- X-DSPAM-Result : Spam
- X-DSPAM-Processed : Sat Jan 27 14 :41 :50 2007
- X-DSPAM-Confidence : 0.8531
- X-DSPAM-Probability : 1.0000
- X-DSPAM-Signature : 5,45bb569e311183356076796

Ham

- X-DSPAM-Result : Innocent
- X-DSPAM-Processed : Mon Jan 29 09 :36 :43 2007
- X-DSPAM-Confidence : 0.9982
- X-DSPAM-Probability : 0.0000
- X-DSPAM-Signature : 5,45bdb21b311182003211484

Quelques options

DSPAM est doté aussi de quelques options intéressantes

- Lien direct avec ClamAV
- Avertissement sur l'état de la quarantaine
- Possibilité de créer des groupes
  - globaux (classification <1000 hams ou < 250 spams)
  - mixés

La correction

La correction peut être effectuée par deux moyens

- Le renvoi du mail mal classé
  - En redirigant le message (signature dans une en-tête)
    - Impossible pour Outlook express, Mail de MacOS
    - Ajout d'un plugin pour Thunderbird
  - En faisant suivre (signature en fin de message)
    - Mail changé
    - Signature numérique éventuelle cassée !
- L'utilisation de l'interface web

Les contraintes d'installation

Elles sont nombreuses, car l'outil est extrêmement paramétrable.

- Choix du backend (hashed, MySQL, Oracle, etc)
- Choix des méthodes (bayésien, markovien, etc.)
- Choix des améliorations (élimination du bruit)
- Choix du paramétrage par défaut
- Choix des options laissées aux utilisateurs
- Choix du processus de réapprentissage
  - Création ou non d'une adresse unique
- Constitution de groupe

Les contraintes d'utilisation

Elles sont nombreuses :

- Problème de l'espace utilisé
  - 10 Mo par personne pour les tokens
  - Autant pour les signatures
  - Obligation de nettoyage régulier
- Problème de performances
  - De 0,10 à 2 secondes par mail
  - Optimisation par InnoDB ?

Les autres défauts

- Interface en anglais uniquement
- L'affichage des graphes est consommateur en CPU
- Problème de droits entre dspam et son interface web

A l'UT1

Choix faits

- Utilisation de MySQL, mais "out of the box" (pas d'innoDB)
- Utilisation des processus par défaut
- Latitude totale pour l'utilisateur
- Utilisation de la réduction de bruit
- TOE (Train on Error) par défaut pour économiser de la place
- Utilisation des virtual users pour créer une adresse unique de redirection
- Limitation de la taille pour analyse des mails
- Création d'une base de départ

Quelques Trucs

Choix faits

- Traque des machines spammeuses
  - Utilisation de l'option "TrackSources spam nonspam"
  - On enlève les MX locaux "LocalMX 127.0.0.1 193.49.48.245 193.49.48.250"
- Enlever les en-tête d'autres outils
  - IgnoreHeader X-Spam-Status
- Mettre dans les en-têtes les critères de choix
  - Preference "showFactors=on"

Efficacité

- Taux de réussite de 99,88% en s'astreignant à la correction
- Tout manquement se paie cash !
- Ré-utilise massivement les méthodes antispam classiques
  - Greylisting (ou plutôt sortie du greylisting)
  - Résultat RBL
  - Résultat Dialup
- Et tout ça sans O.C.R. !
- Avantages fonctionnels
  - Responsabilité individuelle
  - Relativement simple
  - Adaptation à l'utilisateur (mail, web)

Conclusion

DSPAM est un outil

- Extrêmement performant,
- Hautement adaptatif,
- Assez facile d'utilisation,
- Mais qui ne tolère par l'approximation
  - De l'administrateur
  - De l'utilisateur

Merci de votre attention

Des questions ?