Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
PREMIER MINISTRE

S . G . D . S . N
Agence nationale de la sécurité des systèmes d'information
CERTA

Paris, le 03 octobre 2005
N° CERTA-2005-INF-004

Affaire suivie par :
CERTA

NOTE D'INFORMATION DU CERTA

Objet : Limiter l'impact du SPAM

Gestion du document

Référence	CERTA-2005-INF-004
Titre	Limiter l'impact du *SPAM*
Date de la première version	03 octobre 2005
Date de la dernière version	13 juillet 2006
Source(s)	
Pièce(s) jointe(s)	Aucune

TAB. 1 – *Gestion du document*

Une gestion de version détaillée se trouve à la fin de ce document.

1 Définitions

Tout internaute reçoit dans une plus ou moins grande mesure des courriers électroniques qu'il ne souhaite pas recevoir. En général, il s'agit de courriers publicitaires (diverses formes de pornographie, médicaments, papiers officiels, crédits, ...), mais il peut aussi s'agir de propagande ou de diverses formes d'escroquerie.

Le *SPAM* désigne cette forme de courrier. Il est difficile de le définir précisément de façon objective, ce qui constitue une des limites de la lutte contre sa prolifération. Plusieurs définitions existent comme par exemple :

courrier publicitaire non sollicité cette définition ne convient pas tout à fait dans la mesure où, d'une part, ce type de courrier ne manifeste pas obligatoirement une offre commerciale et que d'autre part, les émetteurs du courrier pourront arguer que le courrier est sollicité (arguments souvent invoqués : si un internaute publie son adresse mél, ce serait pour recevoir du courrier ; lorsque l'internaute remplit un formulaire, il est informé qu'il recevra des publicités, ...)[1] ;

pourriel ce néologisme mélangeant « poubelle » et « courriel » est explicite mais il n'aide pas à définir des critères dans la lutte contre ce phénomène ;

courrier illégitime cette expression suggère que le courrier ne se fonde pas sur le bon droit, la raison, etc. ... Cette définition laisse la porte ouverte au spammeur pour expliquer en quoi son message est justifié en droit.

1. Le CERTA ne se prononce pas sur la validité légale de tels arguments, mais constate uniquement que ce sont des arguments employés pour convaincre qu'un courriel n'est pas un courrier publicitaire *non sollicité*.

Globalement le *SPAM* est indésirable parce qu'il engendre du trafic non désiré, le volume de courriers électroniques concerné peut avoir une influence sur le bon fonctionnement de l'Internet et parce qu'il peut porter d'autres formes de malveillances informatiques (*WEB bug*, arnaque, virus, ...).
Dans ce document, *le* SPAM *désigne tout courrier jugé indésirable par son destinataire*.

La lutte contre le *SPAM* conduit à

- qualifier un message : « le message est-il un courrier indésirable ? »
- prendre une décision : accepter d'acheminer, rejeter ou marquer le message.

Dans ce document, on appelle :

- « faux positif » la qualification de *SPAM* attribuée à tort à un message que le destinataire aurait considéré comme acceptable ;
- « faux négatif » la qualification de *non SPAM* attribuée à tort à un message jugé indésirérable par son destinataire.

2 Importance du phénomène

2.1 Origine du *SPAM*

2.1.1 Publicités

Il arrive, fort rarement, qu'un commerçant de bonne foi, mais peu au fait des usages sur l'Internet, inonde ses prospects de courriels publicitaires.
Il arrive plus souvent que des commerçants de bonne foi louent l'usage de listes d'adresses de méls dont la collecte ne respecte pas les règles minimales de déontologie voire de légalité.
Comme la source de la collecte des adresses de méls n'est jamais présentée lors de l'envoi de courriels publicitaires, l'internaute destinataire peut, de bonne de foi, ne pas faire le lien entre le consentement qu'il a pu donner et l'usage qui est fait de son adresse.

Toutefois, bien souvent, le *SPAM* est un courrier publicitaire pour un produit dont la vente est encadrée voire interdite (pornographie, médicament, placement, crédit, diplôme, papiers d'identité, jeu d'argent, contrefaçons de logiciel, ...). Le spammeur dans ce cas est conscient qu'il est, au mieux, aux limites de la loi.

2.1.2 Arnaques

Le *SPAM* peut prendre la forme de diverses escroqueries dont une des plus communes est celle dite *escroquerie nigérianne* ou *fraude 419*. Dans ce type d'arnaque, l'émetteur du courriel se présente comme l'héritier d'un riche notable, parfois dans un pays africain, récemment décédé. Le soi-disant héritier prétend avoir des difficultés pour faire valoir ses droits et propose à la victime d'utiliser le compte en banque de cette dernière et lui propose en échange une rémunération importante pour la gêne occasionnée, d'autant plus qu'à cause des difficultés alléguées, cette dernière doit avancer les frais relatifs à la transaction.
Ce sont invariablement des tentatives d'escroquerie, souvent rendues crédibles par la référence à l'actualité relatif au décès d'une célébrité et à la production de faux documents.

Une autre source d'arnaques exploitant l'actualité survient à l'occasion d'événements majeurs (élection d'un nouveau président par exemple) et à chaque fois qu'une catastrophe naturelle se produit. Dans ce dernier cas, ces arnaques imitent les appels à la solidarité avec les victimes lancés par des associations humanitaires respectables.

2.1.3 Hameçonnage

Le filoutage (ou *phishing*) est la technique qui consiste à envoyer un faux courrier électronique semblant être émis par une institution (banque, site d'enchère en ligne, fournisseur d'accès à Internet, etc. ...) et invitant sous divers prétextes à suivre un lien conduisant sur une copie du site, avec accès protégé, de l'institution visée. L'objectif est de voler les codes d'accès pour accéder au vrai site WEB de l'institution (par exemple, voler le code d'accès d'un client d'un site de banque en ligne, dans le but de se connecter à sa place et procéder à des virements bancaires).

2.1.4 Bombardement de courriel

Ce sont des courriers électroniques dont le contenu n'a pas d'intérêt particulier pour l'émetteur, mais dont l'émission a pour unique but un déni de service sur le destinataire. Le déni de service peut être produit de diverses manières :

Message de grande taille Ce sont des messages qui ont une pièce jointe d'une taille considérable. Au dela d'une certaine taille de message, les serveurs de messagerie et les équipements de filtrage peuvent rencontrer des problèmes de traitement ou de stockage temporaire.

Grand nombre de messages Le nombre de messages envoyés par unité temps peut aussi être la caractéristique d'une attaque en déni de service.

Ce type d'attaque engendre un effet d'amplification par le nombre de messages d'erreur potentiels (DSN)[2]. En cas d'attaque, il est recommandé de désactiver la fonction d'émission de messages d'erreur.

Grand nombre de destinataires L'attaque consiste à noyer le serveur de messagerie sous un grand nombre de messages envoyés à des adresses prises au hasard. L'attaquant ne connait pas les noms des destinataires, mais il essaie de les deviner, par exemple avec des combinaisons fréquentes de noms et de prénoms.

Usurpation du domaine émetteur (*Joe Job*) Le spammeur se fait passer pour un tiers lorqu'il envoie sa campagne de courriel. De nombreux messages vont être refusés et faire l'objet de message d'erreurs.[3] Les messages d'erreurs ne sont pas envoyés au spammeur mais au domaine dont il a usurpé l'identité.

Le trafic de messages d'erreur est une nuisance pour le domaine dont le nom a été usurpé. Cette nuisance se traduit aussi bien par une exploitation des ressources du serveur de messagerie pouvant aller jusqu'au déni de service et par une atteinte à l'image de marque du domaine usurpé (à cause, par exemple, de la nature des produits dont le *SPAM* fait la promotion).

2.2 Étendue de la nuisance

Par nature, le *SPAM* a toujours été considéré comme une nuisance. Les internautes constatent néanmoins une recrudescence du phénomène.

Le phénomène devient de plus en plus une préoccupation pour les opérateurs qui mettent en œuvre l'infrastructure qui délivre le courrier électronique. Ces opérateurs ne sont pas indifférents à devoir acheminer et stocker temporairement du courrier qui en définitive ne sera pas lu par leur destinataire.

Bien qu'il n'y ait pas d'étude définitive sur ce thème, il est généralement reconnu que le *SPAM* est une activité économique rentable. Le coût pour envoyer un message à un nombre massif de destinataires est considéré comme marginal pour l'émetteur, par rapport à la télécopie ou au courrier. La part principale du coût est prise en charge par l'infrastructure et le destinataire.

Devant l'ampleur des vagues de messages considérées comme abusives, une offre d'outils s'est développée. En réaction, les pourvoyeurs de *SPAM* se sont adaptés pour contourner les protections mises en œuvre par ces outils.

2.3 Portée de la lutte contre le *SPAM*

La situation actuelle est que la lutte contre le *SPAM* est facilitée par les comportements et des outils en perpétuelle évolution pour s'adapter à l'évolution des techniques des spammeurs.

Il n'y a pas, actuellement et dans un avenir envisageable, de techniques absolues de lutte contre le SPAM. Il existe cependant un faisceau de techniques qui, lorsqu'elles sont utilisées ensemble, contribuent à diminuer considérablement l'impact du *SPAM*.

3 L'acheminement du courrier sur l'Internet

Le *SPAM* est une nuisance reposant sur un usage abusif du courrier électronique. Pour bien comprendre ce phénomène et les techniques pour s'en protéger, il peut être utile de rappeler le fonctionnement du courrier électronique sur l'Internet.

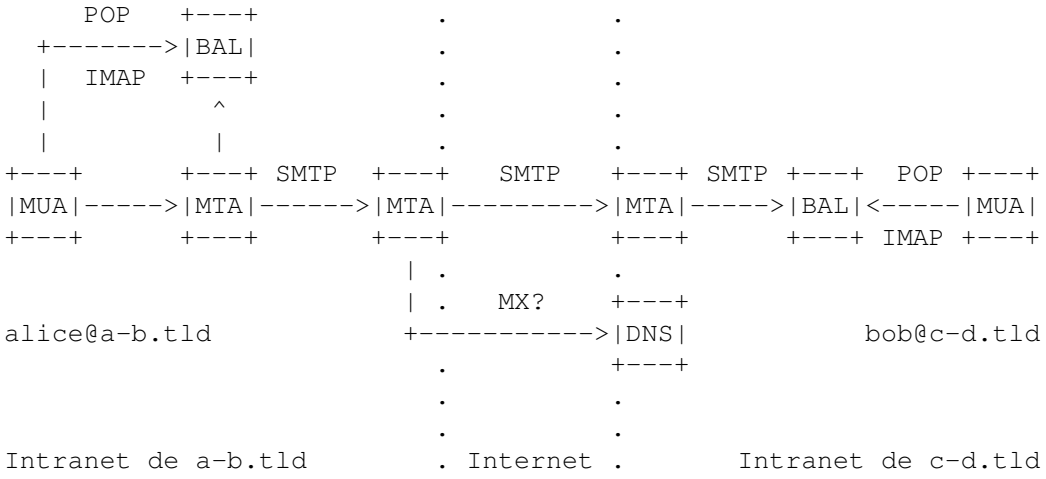
2. cf. section 3.5.
3. cf. section 3.5.

3.1 MTA — MUA

Le courrier électronique sur l'Internet est transmis au travers du protocole SMTP et ses évolutions. La mise en œuvre de ce protocole fait appel à deux types de logiciel :

le MUA *Mail User Agent* est l'outil utilisé pour saisir, lire, sauvegarder des courriers électroniques ;[4]
le MTA *Mail Transfert Agent* est le logiciel chargé de l'acheminement du courrier électronique.[5]

Le transfert du courrier entre deux internautes se passe de la façon suivante.



Alice utilise un logiciel client de messagerie, un MUA. C'est le programme qui lui permet de lire, d'écrire et d'envoyer des courriers électroniques. Alice écrit un message pour Bob. Au moment d'envoyer ce message, le MUA d'Alice contacte le MTA local.

Le rôle d'un MTA est d'acheminer le courriel, en l'aiguillant dans la bonne direction ou d'avertir si le courriel n'a pas pu, pour une raison ou une autre, être délivré à son destinataire.

Lorsque le MUA contacte le MTA, il doit se passer deux choses :

- le MUA doit se présenter auprès du MTA et recevoir l'autorisation d'envoyer du courrier ;
- le MUA doit décrire au MTA à qui le message est destiné et transmettre au MTA le contenu du message.

Dans notre exemple, lorsqu'Alice désire envoyer son message, le MUA d'Alice se connecte au MTA local. Le MTA local lit l'adresse du destinataire du courriel et découvre qu'il ne s'agit pas d'une adresse locale pour laquelle il suffit de déposer le courriel dans une boîte aux lettres (BAL). Le destinataire du courriel est hors du domaine local. Le MTA local aiguille le courriel vers le MTA de l'entreprise.

Le MTA de l'entreprise découvre que le courriel est destiné à une autre entreprise. Le MTA de l'entreprise `a-b.tld` cherche à savoir quel est le MTA de l'entreprise `c-d.tld` (qui est le serveur de messagerie qui accepte de recevoir du courrier pour ce domaine ?). Pour cela il utilise le protocole DNS pour demander le champ `MX`[6] du domaine `c-d.tld`. Avec la réponse du DNS, le MTA de `a-b.tld` peut s'adresser au MTA de `c-d.tld` pour lui transmettre le courrier.

Le MTA de `c-d.tld` reçoit le message et découvre que celui-ci est destiné à Bob et achemine le message électronique vers le serveur de messagerie le plus proche de ce dernier. Arrivé dans le bon service, le message est délivré dans la boîte aux lettres de Bob. Bob consulte son courriel en utilisant un protocole de consultation des boîtes aux lettres, comme par exemple POP3 ou IMAP.

Le protocole SMTP et son extension ESMTP sont utilisés pour la communication entre le MUA et le MTA puis de MTA en MTA.

4. Des logiciels comme `Outlook` ou `Thunderbird`, par exemple, sont des MUAs.
5. Des logiciels comme `postfix`, `sendmail` ou `Exchange`, par exemple, sont des MTAs.
6. cf. section 3.6.

3.2 Relais de messagerie

Un client de messagerie ne s'adresse pratiquement jamais au client de messagerie. En général, le message est convoyé de serveur de messagerie en serveur de messagerie. Un serveur de messagerie qui accepte un courrier pour le transmettre à un autre serveur de messagerie est appelé un *relais de messagerie*.

Normalement un relais de messagerie est conçu pour acheminer le message à l'interieur d'une organisation, ou de l'intérieur de celle-ci vers l'extérieur ou de l'extérieur vers l'intérieur.

On dit qu'un serveur de messagerie est *ouvert* lorsqu'il accepte de relayer des messages issus de n'importe quel domaine (ou plus exactement n'importe quelle plage d'adresses IP) vers n'importe quel domaine (par exemple un serveur de messagerie ouvert en `.gouv.fr` accepterait du courrier venant des plages d'adresses IP de `wanadoo.fr` pour l'envoyer vers `hotmail.com`).

3.3 Le protocole SMTP

Le protocole SMTP est décrit dans le document RFC 821. Cette note d'information présente une vision très simplifiée de ce protocole. Il s'agit d'illustrer les points nécessaires à la compréhension de la lutte contre le *SPAM*.

3.3.1 Transaction SMTP

MTA de a-b.com		**MTA de c-d.com**	
	`EHLO a-b.com` `------->`		le MTA de a-b.com s'identifie

	`250-c-d.com`
	`250-PIPELINING`
	`250-SIZE 10240000`
	`250-VRFY`
	`250-ETRN`
	`250-STARTTLS`
	`250 8BITMIME`
	`<-------`
			le MTA de c-d.com répond

	`MAIL FROM: alice@a-b.com` `------->`		le MTA de a-b.com précise qui est l'émetteur déclaré
	`250 Ok` `<-------`		
	`RCPT TO: bob@c-d.com` `------->`		qui est le destinataire
	`250 Ok` `<-------`		
	`DATA` `------->`		
	`354 End data with <CR><LF>.<CR><LF>` `<-------`		

	`Date: Wed, 28 Sep 2005 11:31:06 +0200`
	`From: Alice <alice@a-b.com>`
	`To: Bob <bob@c-d.com>`
	`Subject: Important!`
	`Message-Id: <20050928113106.5b42ecd6.alice@a-b.com>`
	`Organization: a-b`
	`Mime-Version: 1.0`
	`Content-Type: text/plain; charset=US-ASCII`
	`Content-Transfer-Encoding: 7bit`
	
	`Le texte du message`
	`.`
	`------->`
			le corps du message

	`250 Ok: queued as EC967A4C09` `<-------`		le message est accepté

```
QUIT
------->

221 Bye
<-------
```

En particulier, pour toutes les commandes du protocole SMTP, il existe des codes d'erreurs destinés à l'émetteur.

2xx les codes commençant par 2 sont des codes montrant que l'opération s'est déroulée correctement ;

3xx les codes commençant par 3 signalent des erreurs temporaires ;

...

3.3.2 Quelques autres commandes SMTP

Deux autres commandes SMTP ont une influence notable sur le *SPAM*. Ce sont :

VRFY La commande `VERIFY` sert à vérifier la présence d'un utilisateur ou d'une boîte aux lettres ;

EXPN La commande `EXPAND` sert à vérifier la présence d'une liste de diffusion et d'obtenir la liste de tous les abonnés de la liste.

Ces deux commandes permettent à un spammeur de vérifier ou de confirmer des adresses de courriels.

3.4 L'en-tête d'un message

C'est l'information qui est attachée au message et qui précise les conditions de son acheminement

```
Return-Path: <alice@a-b.tld>
X-Original-To: bob
Delivered-To: bob@c-d.tld
Received: from a-b.tld (unknown [192.168.100.100])
by c-d.tld (Postfix) with SMTP id D2FBEA4C09
for <bob>; Wed, 28 Sep 2005 11:37:10 +0200 (CEST)
Date: Wed, 28 Sep 2005 11:31:06 +0200
From: Alice <alice@a-b.tld>
To: bob@c-d.tld
Subject: Important !
Message-Id: <20050928113106.5b42ecd6.alice@a-b.tld>
Organization: CERTA
X-Mailer: Sylpheed version 1.0.4 (GTK+ 1.2.10; i386-pc-linux-gnu)
Mime-Version: 1.0
Content-Type: text/plain; charset=US-ASCII
Content-Transfer-Encoding: 7bit
Status:

Le texte du message
```

3.4.1 L'en-tête `Reply-To:`

Normalement quand le destinataire souhaite répondre à un message, il utilise la valeur donnée dans le champ `From:` de l'en-tête. Cependant, dans certains cas on peut souhaiter que la réponse ne soit pas acheminée à l'émetteur. Pour cela, le champ `Reply-To:` de l'en-tête permet de spécifier le ou les destinataires de la réponse au courriel.

3.5 Que se passe-t-il en cas de non délivrance d'un message

Si, pour une raison quelconque, le courriel ne peut pas être acheminé jusqu'à son destinataire, une bonne pratique, appuyée par une norme RFC, veut que le MTA retourne un message d'erreur sous la forme d'un courriel qui explique pourquoi le message n'a pas pu être acheminé.

Ce type de message d'erreur est appelé DSN, *Delivery Status Notification*. C'est très utile pour :

- d'un point de vue technique, comprendre les erreurs dans une architecture ou dans la configuration d'un système de messagerie ;
- pour informer l'émetteur que le destinataire n'a pas eu la possibilité de lire le courriel qui lui était destiné.

Pour éviter des boucles (un courriel de d'erreur en réponse à un courriel d'erreur), souvent le courriel d'erreur peut ne pas avoir d'adresse de retour (cf. RFC 821) : `MAIL FROM: <>`.

3.6 Le `MX` Record

Le champ `MX` (abréviation de *Mail eXchange*), identifie dans le DNS le ou les serveurs responsables du traitement des messages pour un domaine particulier.

Par exemple, le `MX` du domaine `certa.ssi.gouv.fr` est la machine `mail.certa.ssi.gouv.fr`. Un mail envoyé à l'adresse de mél `certa-svp@certa.ssi.gouv.fr` est donc en pratique reçu par le serveur `mail.certa.ssi.gouv.fr`.

4 Techniques de lutte contre le *SPAM*

Ce document ne s'intéresse qu'aux mesures techniques et organisationnelles de lutte contre le *SPAM*. Les aspects légaux ne sont pas abordés.

4.1 Techniques basées sur la limitation des ressources

L'idée sous-jacente à ces méthodes est que le spammeur ou le bombardier fait un usage abusif des ressources du serveur de messagerie. En limitant les ressources accordées au transport estimé abusif du courriel on limite l'impact du *SPAM* avec un effet négatif marginal sur le courriel légitime.

Les ressources que l'on peut limiter sont par exemple, selon les options de configuration des outils :

- le nombre de destinataires par message ;
- le nombre de messages par source et par unité de temps ;
- la taille maximale d'un message ;
- ...

4.2 Techniques basées sur la qualification de la source du courriel

L'idée qui est à la base de ces techniques est que le *SPAM* est émis par un serveur.

4.2.1 Existence du domaine émetteur

Certains logiciels de serveur de messagerie ont une option qui permet de vérifier si le domaine émetteur existe. Cela évite le courrier avec un domaine émetteur usurpé.

Le risque de faux positif est limité, en effet (en l'absence de `Reply-To`) il est difficile de répondre à un message pour un domaine qui n'existe pas.

4.2.2 Liste noire

Le principe est de qualifier le courrier en fonction de la réputation du serveur qui l'a émis.

La réputation d'un serveur de courrier qui a émis du *SPAM* récemment est entachée, dans la mesure où l'on suppose qu'il pourrait à nouveau en émettre. Le serveur émetteur est identifié par la seule information fiable : son adresse IP.

1° Une première sorte de liste noire identifie, sur la base de signalement/dénonciation, des serveurs ayant envoyé du *SPAM*.

Le spammeur dont l'adresse IP a été mise en liste noire prend le risque qu'un nombre significatif des messages qu'il envoie soit automatiquement qualifiés de *SPAM* et en définitive jamais lus.

Cette première sorte de liste noire a un intérêt limité par le fait que le spammeur s'adapte et fait en sorte de ne pas être connus sous son adresse IP.

2° Certaines listes noires identifient les relais de messagerie ouverts.
En effet une technique utilisée par les spammeurs est d'identifier des serveurs de messagerie configurés en relais ouverts et de les utiliser pour relayer le *SPAM*. L'adresse vue par le récepteur du courriel est celle du relais de messagerie et non l'adresse du spammeur.
Le parti pris d'une telle liste est que la probabilité de recevoir du *SPAM* d'un relais de messagerie ouvert est plus grande que la probabilité de recevoir un *SPAM* d'un serveur qui ne serait pas ouvert.

3° Certaines recommandations sur la lutte contre le *SPAM* recommandent de refuser le courrier venant de connexions du type de celle qui sont vendues au particulier.
L'idée sous-jacente est que, normalement, un utilisateur particulier de l'Internet, lorsqu'il souhaite envoyer un message configure son MUA pour se connecter directement au MTA du fournisseur d'accès. Un robot d'envoi de courriel, comme certains de ceux utilisés par les spammeurs, s'adresse directement au MTA du destinataire.
Ainsi, une technique de plus en plus fréquemment utilisée par les spammeurs est de prendre le contrôle d'une machine mal protégée, à l'aide d'un cheval de Troie par exemple, et d'y installer un serveur ou un relais de messagerie. Il est en général reconnu que de nombreux particuliers ne configurent pas leur ordinateur domestique pour lui permettre de resister à de telle prises de contrôle. Les machines connectées dans les plages d'adresses IP allouées par les fournisseurs d'accès à leurs clients particuliers sont donc par rapport aux adresses professionnelles plus souvent utilisées pour envoyer du *SPAM*.
Par ailleurs, les adresses allouées aux particuliers sont souvent des adresses dynamiques (l'utilisateur n'a pas toujours la même adresse IP). Il est donc difficile pour un particulier de placer son propre MTA sur ce type d'adresse. Ceci est encore un argument pour dire que, normalement, sur une plage d'adresses IP allouées à la clientèle des particuliers normalement, il n'y a pas de serveur de messagerie sciemment installé par le propriétaire de la machine.
Lorsqu'un courriel provient d'une telle adresse IP, il est plausible que ce soit un *SPAM*.

L'inconvénient est le risque de faux positif pouvant toucher un émetteur de courrier qui a les compétences pour installer un MTA dans de telles conditions défavorables.

4.2.3 Comportement du serveur de messagerie émetteur du courriel

Interprétation des codes d'erreur Le spammeur, dans une démarche pragmatique, peut être tenté de simplifier l'usage qu'il fait du protocole SMTP. En particulier, sachant que la liste d'adresses de mél dont il dispose est imparfaite et dans le but de toucher le plus rapidement possible les destinataires dont les adresses sont valides, il est parfois tenté de ne pas tenir compte des messages d'erreurs émis par le serveur de messagerie à qui il envoie des messages électroniques.

Une technique de lutte contre le SPAM appelée *liste grise* consiste à envoyer un message d'erreur temporaire[7] pour obliger le MTA émetteur à tenter de réémettre le courriel. Si le courrier est effectivement réémis, l'adresse de l'émetteur est mémorisée pendant quelques semaines.

L'idée sous-jacente est qu'un MTA légitime normalement configuré va essayer de réémettre le courriel. En revanche, un logiciel d'envoi massif de messages électroniques ne prendra pas cette peine.

Interprétation du protocole SMTP L'hypothèse qui est faite est qu'un logiciel d'envoi de *SPAM* est fait pour envoyer du courrier et non pour en recevoir. Il sait produire des commandes SMTP, mais ne sait pas les interpréter. Pour savoir si on est en présence d'un serveur de *SPAM* ou d'un serveur de messagerie, on va le tester avec des commandes SMTP. Des exemples de tests :

- vérifier si le serveur émetteur du courriel accepte `MAIL FROM: <>` (accepte-t-il de relayer un message d'erreur) ?
- vérifier si le serveur émetteur accepte de recevoir des courriels sur son adresse de mél `postmaster` : (`RCPT TO: postmaster@emetteur.com`)
- vérifier si le serveur émetteur accepte de recevoir des courriels adressés à l'adresse mél de l'émetteur (`RCPT TO: emetteur@emetteur.com`).

7. c'est à dire un code du type `3xx`.

L'inconvénient est que si l'adresse source du message est falsifiée, comme dans le cas des *joe jobs*, alors le serveur testé n'est pas celui qui a réellement envoyé le courriel.

Réponse à un défi Lorsque le message arrive, il est placé dans une file d'attente. Le serveur de messagerie envoie une demande à l'émetteur pour qu'il s'authentifie. L'authentification consiste en général à donner un mot de passe ou à visiter une page. Cela demande à l'émetteur une phase humaine d'interprétation de la requête.

L'inconvient est que cela ralentit l'acheminement du courrier. Il y a un risque de faux positif.

4.3 Techniques basées sur la qualification du contenu du message

L'idée à la base de ces techniques est que le phénomène du *SPAM* recouvre des contenus assez stéréotypés. L'étude du contenu des courriels peut aider à qualifier en *SPAM* ou non.

4.3.1 Filtre à mots clés

Il s'agit de qualifier de *SPAM* des courriels qui contiennent certains mots clés.

Ce type de technique est insuffisante, il est très aisé pour un spammeur de faire modification mineure à son texte, qui le laisse intelligible mais qui contourne les filtres à mots clés.

Par exemple, à partir du mot CERTA, on peut écrire les variantes suivantes C E R TA, CE.R.TA, C3RTA, C.ERT4, ... Visuellement ces mots restent très proches du mot recherché mais sont différents. Là ou un humain reconnait avec un petit effort quel est le mot maquillé, un programme de reconnaissance automatique peut avoir beaucoup de difficultés.

Les filtres à mots clés sont généralement déconseillés.

4.3.2 Filtre à empreinte

Un filtre à empreinte calcule une signature du contenu d'un courriel et le compare à une base de données d'empreinte de messages considérés comme du *SPAM*.

L'idée sous jacente est que le *SPAM* consiste en l'envoi massif de messages tous identiques.

Il y a deux problèmes derrière ce type d'outils :

1° Comme tous système à signature, un filtre à empreinte ne détecte que les *SPAM* connus.

2° L'autre problème est qu'une modification minime du corps du texte (quelques caractères aléatoires) suffisent pour que la prise d'empreinte soit rendue inefficace puisque chaque empreinte sera différente.

4.3.3 Filtre heuristique

Ces filtres cherchent à établir une probabilité que le message soit un *SPAM* en étudiant son contenu, et le comparant ce dernier avec des caractéristiques de *SPAM* émis dans le passé :

- HTML dans le corps du message ;
- de nombreux mots écrits uniquement avec des lettres majuscules ;
- mots clés correspondants à des produits souvent vantés au travers du *SPAM* ;
- très grand nombre de destinataires ;
- ...

4.4 Techniques basées sur la configuration des systèmes

4.4.1 Politique de lutte contre les logiciels malveillants

De plus en plus de logiciels malveillants (virus, chevaux de Troie, *bot*, ...) installent un serveur de messagerie sur la machine qu'ils ont compromise. Cette fonctionalité des outils malveillants est destinée à faciliter la propagation du *SPAM*.

Lutter contre le *SPAM*, c'est donc aussi lutter contre les logiciels malveillants. Le CERTA a émis de nombreux documents sur la façon de se protéger contre les logiciels malveillants.

4.4.2 Configurer les serveurs de messagerie

Faire une déclaration `SPF` Une façon de contribuer à qualifier un courriel reçu pourrait être « est-ce que le serveur qui l'a émis est une machine reconnue par le maître du domaine émetteur apparent comme légitime pour envoyer du courriel ? ».

Nous l'avons vu, le DNS d'un domaine au travers de son champ `MX` peut indiquer quel est le serveur de messagerie prévu pour *recevoir* le courrier électronique pour le domaine. C'est absolument nécessaire pour envoyer du courrier puisque le MUA ou le MTA doit déterminer quel serveur assure le relayage du message vers le destinataire.

Le champ `MX` ne répond pas à la question qui consiste à déterminer quelles sont les machines prévue pour envoyer du courrier. Il faudrait une sorte de `MX` à l'envers.

C'est l'objet du champ `SPF`. Il sert à désigner dans le DNS les machines autorisées par le domaine à émettre du courrier. Les MTA ont ensuite la possibilité, au moment de la réception de la commande `EHLO` ou `MAIL FROM`, de vérifier si l'adresse IP de l'émetteur du courriel déclarant provenir de `mondomaine.com` est bien une des adresses déclarées par le gestionnaire du domaine `mondomaine.com` comme une adresse autorisée à envoyer du courriel.

L'inconvénient est qu'à l'heure actuelle le bénéfice de l'intérogation du champ `SPF` est faible puisque peu de domaines l'ont rempli.

Un autre inconvénient est que l'interprétation du `SPF` peut s'avérer fausse dans certains cas où l'utilisateur fait suivre son courrier.

Enfin, le `SPF` en tant que tel n'est pas suffisant. Un spammeur pourrait très bien s'acheter ses noms de domaines et déclarer des champs `SPF` pour ses domaines. Les promoteurs de `SPF` mettent en avant la nécessité de la gestion de la réputation.

Éviter les relais non maîtrisés A moins d'avoir une bonne raison de faire autrement, il est conseillé de ne pas laisser un relais de messagerie ouvert sur l'Internet. Un serveur de messagerie raisonnablement configuré ne devrait accepter de relayer que vers les quelques domaines nécessaires. [8]

4.4.3 Détecter les serveurs de messagerie

Normalement le plan de déploiement du parc informatique devrait préciser quelles sont les machines destinées à servir de relais de messagerie.

Si c'est autorisé, scanner le réseau pour découvrir le port 25/TCP ouvert.

4.4.4 Maîtriser ses routes

Certains fournisseurs d'accès offrent un service relatif aux protocoles de routage, tel que BGP par exemple. Les spammeurs ont dès lors la possibilité d'identifier des plages d'adresses IP pour lesquelles aucune route n'est définie, puis de déclarer des routes vers ces plages et d'émettre du *SPAM* depuis ces plages.

Il existe des listes de plages d'adresses IP non attribuées. Aucun trafic ne devrait venir de celles-ci.

Une bonne pratique veut que l'on filtre en entrée et en sortie de son réseau ces plages d'adresses.

4.5 Techniques basées sur le comportement de l'internaute

4.5.1 Utiliser des adresses volatiles

Certains outils d'envoi de *SPAM* collectent les adresses de méls qu'ils trouvent sur la machine compromise sur laquelle ils sont installés. En particulier, un tel logiciel peut chercher dans les boîtes aux lettres. Un abonné à une ou plusieurs listes de diffusion, serait une cible idéale pour de tels outils parce qu'une grosse liste de diffusion avec plusieurs publications par jour est une source de nombreuses adresses de méls sauvegardées sur de nombreuses machines d'internautes.

Pour limiter les risques qu'une adresse de mél soit collectée dans de telles conditions, il est recommandé d'utiliser des adresses jetables pour poster dans les listes de diffusion ou les groupes de *news*, dans un *blog* ou un forum.

Il existe sur l'Internet de nombreux prestataires qui offrent des boîtes aux lettres gratuites. Il ne faut pas hésiter à créer de nombreuses boîtes aux lettres (une par usage).

8. Ce n'est en aucun cas une limitation sur la nature des messages que l'on peut recevoir ou envoyer. C'est une limitation sur l'usage que des tiers pourrait faire du serveur pour envoyer du courriel à d'autres tiers.

4.5.2 Ne pas donner l'adresse de quelqu'un d'autre

Chaque internaute devrait choisir lui-même l'usage qu'il fait de ses adresses de messagerie.

Il faut éviter de donner une adresse de mél d'un tiers dans un formulaire que l'on remplirait à sa place sans son accord préalable, ou en donnant son adresse dans le champ d'un formulaire de site WEB du style « envoyer à un ami ».

Une façon de protéger l'adresse de mél des tiers qui ont fait confiance en vous la donnant, est de mettre les destinataires des courriels dans les champs invisibles tels que Bcc, plutôt que les champs To et CC. C'est important dès lors que l'on a beaucoup de destinataires dans un message.

4.5.3 Éviter la moisson d'adresses de mél

Il existe des outils pour collecter les adresses de mél sans se préoccuper du consentement de leur propriétaire. Il s'agit en particulier d'outils qui extraient les adresses de mél publiées sur les sites WEB. Afin d'éviter que les adresses ne soient collectées de la sorte, il convient d'éviter de publier les adresses sur les sites WEB.

Il est préférable de ne publier que les adresses fonctionnelles plutôt que les adresses personnelles.

Il y a des astuces pour publier les adresses de mél dans l'espoir qu'elles échappent au outils de collecte :

- remplacer le texte de l'adresse par une image qui contient ce texte ;
- utiliser une devinette qui oblige un être humain à extraire l'adresse. Par exemple

  `certa123SPAM456-svp@certa.ssi.gouv.fr` retirer les chiffres et le mot *SPAM* pour avoir l'adresse valide

L'inconvénient de ce type d'astuce est qu'il contribue à altérer l'accessibilité de la page WEB.[9]

4.5.4 Ne *jamais* répondre aux *SPAM*

Répondre à un *SPAM*, c'est confirmer au spammeur que l'adresse à laquelle il a envoyé un message existe vraiment. Indirectement, c'est s'exposer à recevoir beaucoup d'autres *SPAM*.

Ne pas répondre ne suffit pas. Encore faut-il configurer son lecteur de messagerie pour qu'il ne réponde pas à la place du lecteur :

- ne pas envoyer d'accusé de réception ;
- interdire au lecteur de charger des images ou tout autre objet sur des sites distants ;
- interdire l'exécution du code javascript;

5 Atténuer les effets de bords de la lutte contre le *SPAM*

> Presque tous les hommes meurent de leurs remèdes, et non pas de leurs maladies.
>
> Molière
> *Le malade imaginaire*

5.1 Les relais de messagerie ouverts

Le fait qu'un serveur de messagerie soit configuré en relais et que ce relais soit ouvert n'est pas un problème de sécurité en tant que tel pour l'organisation qui met en œuvre ce serveur.

La norme qui définit le protocole SMTP, le RFC 821, indique même que le relais est un bien meilleur choix en matière de sécurité que d'avoir à définir explicitement la route par laquelle doit passer le courriel.

Cependant, tant de nombreux abus ayant été commis par des tiers indélicats cherchant à masquer leur origine, qu'un serveur de messagerie configuré en relais ouvert a de grandes chances de se retrouver rapidement dans une liste noire.

La conséquence de cette mise en liste noire est que le courrier émis par l'organisation a une plus grande probabilité d'être qualifié de *SPAM* par les destinataires et donc d'être rejeté.

9. Une image, par exemple devrait être accompagnée d'un texte alternatif pour ceux qui n'ont pas la possibilité de visualiser l'image.

5.2 Les formulaires WEB pour envoyer un courriel

De nombreux sites WEB interactifs offrent la possibilité d'envoyer des messages à l'institution qui met en œuvre le site au travers d'un formulaire. Ce formulaire est ensuite traduit en courrier.

La configuration du formulaire et du serveur de messagerie qui traite les saisies dans ce formulaire, devrait être telle que seule l'institution peut recevoir des courriels saisis dans le formulaire. Le formulaire ne doit pas servir de relais pour envoyer des messages à n'importe qui dans le monde.

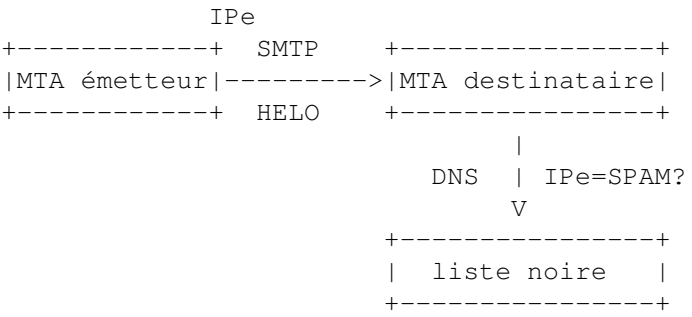
5.3 Point d'accès ouverts ou faiblement protégés

La lutte contre le *SPAM* conduit les spammeurs à rechercher de plus en plus à se cacher pour commettre leur action. Au même titre que la compromission d'ordinateurs peu protégés, l'utilisation de points d'accès ouverts permet d'envoyer du *SPAM* relativement discrètement.

Une organisation qui envisage de fournir des points d'accès doit s'interroger sur le besoin de les laisser ouverts et vérifier que les points d'accès sont configurés conformément à la politique retenue.

5.4 Utiliser une liste noire

Le principe de fonctionnement d'une liste noire est le suivant.



Le serveur de messagerie, lorsqu'il reçoit une connexion SMTP, acquiert l'adresse IP de l'émetteur. Il interroge alors la ou les listes noires auxquelles il est abonné afin qu'elles qualifient ou non cette adresse IP d'émetteur récent de *SPAM*.

Le protocole utilisé pour interroger une base de données d'adresses IP est naturellement le protocole DNS.

La qualité de la qualification dépend du sérieux avec lequel la liste noire est gérée. Si le gestionnaire de la liste noire n'accepte pas rapidement les nouveaux signalements alors le taux de faux négatifs augmente. Si au contraire le gestionnaire de la liste tarde à nettoyer de sa base les adresses des serveurs de messagerie ayant fait l'effort nécessaire pour ne plus propager du spam, alors le risque de faux positifs augmente.

5.5 Utiliser une liste noire gérée par un tiers

Indépendamment de la qualité de la gestion de la liste, utiliser une liste gérée par un tiers suppose que l'on ait confiance dans la discrétion et l'intégrité de ce tiers. En effet, le fonctionnement même de la liste noire fait qu'à chaque réception d'un message l'adresse IP de l'émetteur et/ou le nom de domaine de l'émetteur est envoyé à la liste noire.

Avant d'utiliser une liste noire, il convient de s'interroger sur ce qu'elle recense exactement (IP connues pour avoir émis du *SPAM* ou relais de messagerie) et sur la façon dont elle est gérée. Par ailleurs, il est important de s'interroger régulièrement sur la qualité de la gestion mise en œuvre dans cette liste.

5.6 Administrer sa propre liste noire

Les listes noires de serveurs de messagerie existent depuis longtemps. La plupart d'entre elles ont été des projets bénévoles gérés par des passionnés voulant aider à lutter contre le *SPAM*.

Certains spammeurs ont réagi à l'existence de ces listes par le déni de service. Elles ont été tellement saturées de requêtes que :

- soit elles deviennent inutilisables ;
- soit l'augmentation de la bande passante nécessaire pour que les requêtes légitimes ne soient pas altérées par le *SPAM* ou le déni de service est telle que le coût devient prohibitif pour un bénévole.

Ce problème devient vraiment important si on offre le service de liste noire à des tiers sur l'Internet. Pour un usage interne, le risque est moindre.

5.7 Liste blanche

Dans certains cas, recevoir le courrier légitime de certains émetteurs est une priorité. Pour une raison ou pour une autre, il est possible que cet émetteur se trouve placé dans une liste noire. Mettre une adresse en liste blanche, c'est considérer que quoiqu'il arrive, tout courrier venant de cette adresse sera acheminé.

5.8 Liste grise

Plutôt que d'utiliser une liste noire difficile à gérer, on peut partir du principe suivant : « *a priori* tout message provenant d'une adresse IP inconnue est temporairement rejeté à moins qu'il ne soit réémis » . Toute adresse IP ayant réémis un message est considérée comme utilisant un serveur de messagerie capable d'interpréter les messages d'erreur (d'indisponibilité temporaire), donc il ne s'agit pas d'un automate simplifié d'émission de *SPAM*. L'adresse IP ayant réémis le message obtient alors une bonne réputation qu'elle conserve en étant mise dans une liste blanche pendant une durée limitée de quelques semaines. Ensuite elle doit repasser le test.

5.9 Adresse de désabonnement

Au regard du droit français, les adresses utilisées pour envoyer du courrier à des personnes privées, doivent avoir été au préalable collectées de façon loyale, c'est à dire avec le consentement de leur propriétaire quant à leur utilisation future, à des fins publicitaires par exemple.

Dans l'hypothèse où l'internaute aurait donné son consentement, il se peut qu'à la longue, il ne souhaite plus recevoir ce type de courrier. La loi oblige l'émetteur d'un courrier publicitaire à offrir un moyen de se désabonner. L'utilisation d'une telle adresse de désabonnement avec un émetteur de bonne foi est de nature à limiter le nombre de courriers indésirables.

Cette démarche fonctionne avec les acteurs honnêtes du marketing, qui dans une démarche commerciale bien comprise, ne cherchent pas à mécontenter leurs prospects par des courriers indésirables.

Le spammeur, quant à lui, n'a pas à se préoccuper du mécontentement des destinataires puisque son activité ne lui permet d'espérer la réponse que d'une infime partie des destinataires. Il lui importe plus d'être lu que de contenter. La première condition pour qu'un message soit lu est que l'adresse du destinataire existe. Le spammeur est donc tenté de vérifier que l'adresse est valide.

Une réponse à une adresse de désabonnement indique clairement au spammeur que l'adresse de désabonnement a été extraite, c'est à dire qu'un être humain a lu le courriel et y a répondu. L'adresse de destination du courriel est donc bien l'adresse d'une personne. Le spammeur a donc tout intérêt à réutiliser cette adresse.

Certains thèmes dans le contenu d'un message l'identifie clairement comme du *SPAM*. Il est donc impératif de ne pas répondre à ce type de message.

Toutefois, l'apparence ne permet de qualifier un courrier en *SPAM*. En effet, un spammeur astucieux peut concevoir un courriel ressemblant à un message d'une société commerciale sérieuse ayant une démarche de marketing loyale avec une adresse de désabonnement appartenant au spammeur. [10]

Dans ce cas, la seule chose qui permet de distinguer un message légitime d'une copie est l'adresse IP de l'émetteur. L'apparence du courriel n'est d'aucun secours. L'adresse IP de l'émetteur n'est pas toujours présente dans le message, mais quand elle est présente c'est une donnée disponible dans l'en-tête du message.

Les adresses de désabonnement ne sont donc un outil utile que pour les utilisateurs capables de lire les en-têtes des courriels qu'ils reçoivent.

La différence subtile pour le grand public entre un courrier publicitaire légitime et certains *SPAM*, conduit à ce que l'obligation, sur laquelle il convient de ne pas revenir, faite aux acteurs du marketing de placer une adresse de désabonnement dans les messages publicitaires, est donc en contradiction avec les recommandations faites au grand public de ne pas répondre à un *SPAM*. Seul un internaute avisé et formé à la lecture des en-têtes d'un courriel peut faire la distinction entre un vrai courriel publicitaire (sollicité ou non) et un *SPAM* et utiliser sans risque ce droit qui lui est accordé.

10. C'est précisément ce qui est fait dans les entreprises de hameçonnage.

6 De la bonne utilisation des techniques

L'objectif de la lutte contre le *SPAM* est d'acheminer le courrier légitime avec le minimum de faux positifs et de faux négatifs dans la détection du courrier indésirable. Une seule technique peut s'avérer insuffisante. L'usage de plusieurs techniques correctement maintenues peut conduire à une lutte anti-*SPAM* sinon absolue du moins relativement efficace. L'attention du lecteur est néanmoins attirée sur le fait qu'utiliser de nombreuses techniques peut compliquer la mise au point de l'architecture du système de messagerie ou la recherche d'erreurs.

6.1 Techniques utilisables par les administrateurs de serveurs de messagerie

6.1.1 Filtres bayésiens collectifs

Lorsque le réseau est assez homogène dans les centres d'intérêts de ses utilisateurs, il peut être envisagé de mettre en place une solution centralisée de filtrage.

6.1.2 Filtrage anti-virus

Une politique de lutte contre les virus est utile dans la lutte contre le *SPAM*. Le CERTA a émis de nombreux documents de bonnes pratiques pour lutter contre les virus informatiques.

L'attention du lecteur est attiré sur le fait que les chevaux de Troie qui participent au relayage du message sont souvent des logiciels de type *bot* ou *zombie*, connus pour être relativement mal détectés par les logiciels anti-virus. La lutte contre les virus ne saurait donc se baser uniquement sur les logiciels anti-virus.

6.1.3 `SPF`

La déclaration et l'interrogation des champs `SPF` dans le DNS est une technique qui pourrait à l'avenir aider à mieux qualifier les messages. Les administrateurs systèmes sont invités à déclarer leur champs `SPF`, après avoir vérifié que les usages relatifs au renvoi du courrier en vigueur dans leur domaine ainsi que leur logiciel de messagerie sont compatibles avec `SPF`.

6.2 Techniques utilisables par les utilisateurs de la messagerie

6.2.1 Filtrage de contenu dans le client de messagerie

La plupart des logiciels de messagerie modernes offrent la possibilité de qualifier les messages sur leur contenu. L'utilisateur peut indiquer finement au logiciel ce qu'il considère ou non comme du *SPAM*.

7 Que faire en cas d'incident de *SPAM* ?

7.1 Un traitement d'incident classique

Le *SPAM* est souvent le symptôme de problème de sécurité (mauvaise architecture, mauvaise configuration, compromission de machines, ...). Dans ce cas, il convient de faire un traitement d'incident (cf. note d'information CERTA-2002-INF-002).

7.2 Désactiver la fonction DSN

En cas de réception d'un nombre massif de courriers avec du *SPAM*, il est fort propable que de nombreux destinataires n'existent pas. Le serveur de messagerie destinataire va normalement générer de nombreux messages DSN. En cas de *SPAM* les émetteurs sont souvent fantaisistes, si bien qu'envoyer des messages d'erreur ne fait qu'amplifier le déni de service.

En cas d'attaque, il est donc recommandé de désactiver la fonction d'envoi de messages d'erreur.

8 Quand contacter le CERTA ?

Le nombre d'incidents de *SPAM* est considérable. Le CERTA ne peut pas traiter tous les incidents de *SPAM*.

Le CERTA a la responsabilité d'aider les administrations de l'État à répondre aux incidents de sécurité qui les affectent. Dans ce contexte, le CERTA traite de nombreux incidents de *SPAM* affectant les administrations.

8.1 Quand l'incident affecte une administration

Lorsqu'un incident de *SPAM* (publicité abusive par son volume, bombardement de courriels, usurpation de l'adresse, ...), ou au contraire que les conséquences des mesures de lutte contre le *SPAM* (liste noire, ...), ou enfin que les moyens mis en œuvre pour diffuser le *SPAM* concernent une administration (relais ouvert), le CERTA est compétent pour traiter l'incident de sécurité.

Le *SPAM* peut être le symptôme de problèmes techniques plus profonds comme la compromission de machines, le vol de données personnelles, ... à ce titre le CERTA attache une certaine importance aux machines impliquées dans la diffusion de *SPAM*.

8.2 Quand vous ne savez pas à qui vous adresser

Le CERTA est compétent pour recevoir les rapports d'incidents concernant les incidents de *SPAM*, les incidents liés aux effets de bord de la lutte contre le *SPAM* ou lorsque les moyens utilisés pour se propager affectent la France.

Le CERTA n'a pas toujours la ressource nécessaire pour traiter finement ce type d'incidents, il pourra néanmoins informer les parties concernées en donnant les conseils nécessaires pour que cesse le trouble.

9 Documentation

- Signal Spam : association de type loi 1901 ayant pour objet de fédèrer les efforts de tous pour lutter contre le SPAM.
  http://www.signal-spam.fr
- l'escroquerie nigérianne :
  http://www.minefi.gouv.fr/tracfin/questions.htm#nigerianne
- Les bons réflexes en cas d'intrusion :
  http://www.certa.ssi.gouv.fr/site/CERTA-2002-INF-002
- les routes qui ne devraient jamais apparaître sur l'Internet :
  http://www.cymru.com/Bogons/
- Le comité réseau des Université (CRU), maintient un document sur les aspects pratiques de la lutte anti-spam et en particulier sur la configuration de certains outils :
  http://www.cru.fr/antispam/index.html/doku.php

Gestion détaillée du document

03 octobre 2005 version initiale.

13 mars 2006 quelques compléments, corrections mineures.

13 juillet 2006 lien sur Signal Spam, quelques corrections mineures.