Lutter contre les pourriels

Guide de survie anti-spam, sous GNU-Linux

INDEXATION DU DOCUMENT			
	TITRE : Lutter contre les pourriels		*REFERENCE* :
ACTION	*NOM*	*DATE*	*SIGNATURE*
RÉDIGÉ PAR	Asdrad Torres	25 mai 2004	

SUIVI DU DOCUMENT			
INDICE	DATE	MODIFICATIONS	NOM

Table des matières

Résumé

Cet article présente comment configurer un ordinateur fonctionnant sous une distribution récente de GNU-Linux afin de se protéger du spam. Le but est d'obtenir une vraie protection, efficace et sans préjudice pour l'utilisateur. L'investissement demandé est en rapport avec le résultat.

Guide de survie anti-spam, sous GNU-Linux

1 Objet

Je me propose d'exposer ici comment configurer un poste de travail individuel sous Linux, dans le seul but d'éliminer les pourriels (spam). Il ne s'agit donc pas d'explorer la configuration d'un serveur de courrier, dans le but de protéger des pourriels l'ensemble des utilisateurs de ce serveur. Je me limiterai bien à exposer une solution applicable à un utilisateur unique. Elle sera néanmoins extensible[1] à un poste "familial" sur lequel plusieurs utilisateurs, chacun dotés d'un compte personnel, se succèdent. Dans ce cas, chaque utilisateur peut être considéré, à tout de rôle, comme un utilisateur unique.

Je présente la configuration de mon poste qui tourne sous GNU-Linux Mandrake 10.0 et exploite l'environnement de bureau KDE 3.2. Si de nombreuses informations présentes dans ce document sont extrapolables à d'autres environnements, je suis bien incapable de préciser exactement dans quelle mesure. Chacun devra donc effectuer les ajustements nécessaires.

Vue la longueur de cet article, vous vous dîtes probablement que "ça ne va pas être de la tarte". En fait, vous devrez effectuer de nombreuses opérations mais aucune, prise individuellement, ne vous demandera d'avoir les connaissances d'un spécialiste. Il s'agira plutôt d'une suite de manipulations longue et méticuleuse plutôt que complexe et difficile. Mais je ne vous cache pas que ce sera plus compliqué que de cocher une case "anti-spam", comme le proposent certains logiciels de courrier. Rassurez-vous...

Si la difficulté n'est pas comparable, l'objectif que nous poursuivons n'a rien à voir avec ce que ces gadgets permettent d'obtenir. Je n'ai pas cherché à supprimer 40% des pourriels que je reçois au risque de perdre quelques bons courriels. D'une part, j'ai voulu détecter automatiquement 99,8% des pourriels tout en réduisant le risque de suppression abusive en deçà de 1/10000 ! D'autre part, je ne cherche pas uniquement à me protéger de manière défensive mais aussi à lutter contre les spammeurs, afin de faire valoir mon droit à utiliser l'internet sans être perpétuellement assailli, à l'insu de mon plein gré ;-)

Enfin, une fois l'installation effectuée, j'ai voulu que le dispositif mis en place ne pénalise en rien l'utilisation du courriel. À la différence de nombreux outils anti-spam proposé ici ou là, la configuration que je décris ici n'a produit aucun ralentissement sensible du fonctionnement de mon logiciel de courrier, ni des mes autres applications.

2 Présentation du problème

- Vous disposez d'une ou plusieurs boîtes aux lettres (BAL) qui sont hébergées sur des serveurs : chez votre fournisseur d'accès, chez un fournisseur gratuit ou payant de boîte aux lettres, chez votre association ou votre entreprise.
- Vous relevez et/ou consultez votre courrier à l'aide d'un logiciel spécialisé, en POP ou en IMAP. Personnellement, j'utilise Kmail. Si vous consultez vos courriel à travers le Web (une application du genre webmail), cet article ne vous apportera aucune solution pratique contre les pourriels.
- Vos boîtes sont remplies de courriels non sollicités, rendant difficile et pénible l'utilisation du courrier électronique. Si les pourriels sont rares et peu gênant, la solution la plus appropriée est celle que vous pratiquez déjà : les éliminer à la main.

3 Lutter contre les pourriels

Trois manières d'intervenir

- empêcher un pourriel déjà émis d'échouer dans votre BAL
- éliminer automatiquement les pourriels que vous avez reçus
- empêcher les expéditeurs de pourriels d'émettre leurs saletés

[1] En fait, réplicable.

3.1 Panorama des procédés

3.1.1 Bloquer la diffusion

Une fois qu'un pourriel a été "lancé" sur l'internet, il est parfois possible de l'empêcher d'arriver dans votre BAL. En effet, pour se diffuser, tous les courriers électroniques utilisent des relais de diffusion. Or, certains relais sont connus pour le complaisance à l'égard du pourriel. En refusant systématiquement de communiquer avec ces relais on se prémunit des nuisances qu'il génèrent.

3.1.2 Trier à l'arrivée

Les pourriels qui n'ont pas pu être bloqués pendant leur cheminement, arrivent chez hébergeur de courrier. La seule chose que l'on peut faire alors est de trier les courriers reçus. Ceci peut être fait, partiellement ou pas du tout, par votre hébergeur. Quoiqu'il en soit, vous aurez toujours à faire un tri final puisque vous seul êtes capable de dire si un message est, pour vous, licite.

3.1.3 Empêcher ou dissuader l'émission

Les pourriels que vous avez reçu contiennent de nombreuses informations. Certaines d'entre elles pourront être exploitées pour mieux reconnaître les pourriels, dans les flot de tous les méls. D'autres informations permettront d'identifier l'origine du pourriel. On pourra alors engager une action ciblée en vue de contraindre l'émetteur à cesser ses pratiques.

3.1.4 Coopérer

Dans cette lutte multiformes, l'utilisateur particulier n'est pas seul. La grande majorité des opérateurs de l'internet, qu'ils soient marchands, étatiques, associatifs ou individuels, est consciente du préjudice collectif causé par les pourriels. De plus, certains services coopératifs, marchands ou non, épaulent l'utilisateur dans les opérations complexes de détection automatisée des pourriels et de traque des fauteurs de troubles. Grâce à ces services, tout possesseur d'un ordi relié à l'internet peut, quel que soit son niveau technique, se protéger des pourriels et lutter contre leurs auteurs. C'est, avant tout, une question de détermination.

Les pourriels étant un fléau dont est victime la communauté des internautes, la lutte doit nécessairement se situer à deux niveaux : individuel et collectif.

3.2 L'action individuelle

3.2.1 L'illusion de la rétention

On dit souvent que la première action individuelle envisageable est préventive. Moins vous diffusez vos adresse de mél, moins vous risquez de recevoir des pourriels. L'affichage de vos adresses sur vos pages Web, la participation à des forums publics, l'utilisation de sites marchands sont autant de sources de diffusion de vos adresses. "S'il est irréaliste de ne pas diffuser ses adresses (sinon, à quoi bon utiliser le mél ?), on peut néanmoins les utiliser avec discernement." Telle serait la maxime des tenants de cette doctrine.

Une première idée est d'utiliser plusieurs adresses, en fonction de la publicité que l'on souhaite leur donner :

– une adresse pour participer aux forums grand public,

– une adresse professionnelle publique,

– une adresse professionnelle restreinte,

– une adresse perso pour les amis,

– une adresse perso générale.

J'en suis venu à la conclusion que ces mesures préventives ne sont utiles que si l'on part battu d'avance. Le principe est le suivant : si l'une de ces adresses est assaillie par les spammers, au point de devenir inutilisable, on pourra toujours l'abandonner, sans tout perdre. Autrement dit, on est déjà prêt à abandonner le terrain, avant même d'avoir livré bataille ! Ne vous y trompez pas, il s'agit bien d'un conflit dont l'enjeu est l'occupation d'un territoire.

Si, au contraire, on souhaite pouvoir continuer à utiliser toutes ses adresses, leur multiplication ne fait que compliquer les choses. Une adresse professionnelle, une adresse perso et une adresse anonyme (pseudo) devraient couvrir les besoins de la grande majorité des utilisateurs.

Enfin, n'oublions pas que toutes ces précautions se révéleront vaines dès l'instant où l'un de vos correspondant, adepte ou victime de Windaube, aura été attaqué par un virus. Du jour au lendemain, son carnet d'adresses "Outlook" sera exposé au grand jour, révélant à tous les adresses que vous aviez si parcimonieusement distillées.

3.2.2 Trier, éliminer

La première vraie mesure de lutte individuelle anti-spam est une mesure de protection. Elle consiste à trier automatiquement tous les courriels que vous recevez en deux catégories : les pourriels (spam) et les courriel licites (ham). Si vous relevez votre BAL avec POP, le tri s'effectuera sur votre ordinateur, dans les dossiers où vous rapatriez les messages. On peut donc partiellement l'effectuer sans être connecté à internet. C'est là tout l'avantage du POP. Si vous consultez votre BAL par IMAP, le tri est a effectuer dans la BAL elle-même. Vous devez alors être connecté à internet pendant le tri.

Pour faire ce tri, on utilisera un logiciel spécialisé (`spamassassin`) qui fonctionnera en osmose avec notre logiciel de courrier (`KMail`). Spamassassin analyse les courriels suivant deux familles de critères. L'allure générale du message, les expressions utilisées ou les types de pièces jointes permettent déjà de se faire une première idée. On la complète par une étude plus spécifiquement "réseau", où l'on va évaluer l'existence, la plausibilité ou la réputation des adresses internet figurants dans le courriels. Ce type d'étude suppose l'interrogation de bases de données et donc d'être connecté à internet. En pratique, cette seconde phase est réservée aux utilisateurs disposant d'un accès permanent, pas nécessairement à haut débit.

3.3 L'action collective

Au niveau individuel, la seule mesure efficace se limite donc au tri des courriels que l'on reçoit. L'idéal serait de ne pas en recevoir mais cela dépasse nos seuls moyens de lutte.

La lutte collective se décline sur deux axes. D'une part, on contribuera à l'identification des spammers. D'autre part, on contribuera à réduire au silence les spammers identifiés.

Des **systèmes coopératifs** de lutte anti-spam permettent de repérer les spammers, les relais de réexpédition derrière lesquels ils se dissimulent et les contenus spammés. Transmettez-leur les messages que vous avez **formellement** identifiés comme étant des pourriels, vous améliorerez les capacités de détection de tous. Ce sont eux qui maintiennent à jour les bases de données interrogées par spamassassin. Ils participent donc au blocage ou au filtrage des pourriels existants mais sont sans effet sur l'émission de pourriels.

Une seconde catégorie de systèmes est plus orientée vers la traque des spammers. Ces systèmes fournissent des outils d'analyse des pourriels reçus et permettent d'adresser facilement des plaintes aux opérateurs qui se seront prêtés, à l'insu de leur plein gré ou non, à l'émission du spam. Ils aident les opérateurs concernés en les aidant à identifier les fauteurs de trouble. Si les spammer persistent malgré les mise en gardes, il seront finalement privés d'accès internet. Ces outils visent donc à empêcher l'émission de spam, soit en dissuadant les auteurs de poursuivre leurs pratiques soit en les empêchant matériellement de le faire.

Si le tri automatique est bien fait, tous les messages classés comme *spam* seront vraiment des pourriels. Les messages classés comme *ham* contiendront tout les messages licites et quelques pourriels résiduels. Il est presque impossible d'arriver à un classement parfait. On préférera avoir quelques pourriels classés comme ham plutôt que d'avoir un seul message licite classé comme spam.

4 Dispositif anti-spam

4.1 Présentation du dispositif

Notre dispositif sera composé de :

– un système de filtrage
– un système de reporting
– un système de traque

4.2 Filtrage

4.2.1 L'application principale

Sur Linux-Mandrake, spamassassin est disponible sous forme de paquetage directement installable (rpm). Son installation est donc automatisée.

Sa configuration serait plus problématique sans l'existence d'un site spécialisé dans cet opération. Le site propose un formulaire avec des boutons radio et des cases à cocher. Il génère automatiquement un fichier de configuration et indique où l'utilisateur doit copier ce fichier. Il suffit de se laisser guider...

Mais il restera toujours une partie du réglage qui sera totalement spécifique à chaque utilisateur. En effet, spamassassin est capable personnaliser sa manière de trier en fonction des types de messages que chacun de nous reçoit. Il faudra donc, avant toute utilisation, fournir au programme un paquet, aussi important que possible de courriels licites que l'on a reçu. Si l'on utilise son logiciel de courrier comme outils d'archivage des messages, l'opération est simple puisqu'il suffit d'alimenter le système d'apprentissage avec nos archives[2]. Dans le même ordre d'idées, on pourra lui indiquer un paquet de pourriels dont on se sera préalablement assurés qu'il s'agit indubitablement de pourriels. Lors de cette phase d"éducation de spamassassin, toute erreur d'appréciation de la part de l'utilisateur se répercutera durablement sur les capacité de détection future du programme.

4.2.2 Razor

Razor est une des applications permettant d'interroger et d'alimenter les bases de données coopératives de lutte anti-spam. Si razor est installé, il sera utilisé par spamassassin, à condition qu'on l'ait demandé dans le fichier de configuration de spamassassin. Sinon, spamassassin fonctionne quand même, mais moins bien.

Sur Linux-Mandrake, razor est disponible sous forme de paquetage directement installable (rpm). Son installation est donc automatisée. Il faudra cependant le configurer par les deux commandes suivantes :

```
razor-admin -d -create
razor-admin -register
```

La première commande crée les fichiers de configuration de razor. La seconde crée automatiquement un code d'identification par lequel l'utilisateur sera connu par le serveur razor.

4.2.3 Pyzor

Pyzor est une application du même type que razor. Si pyzor est installé, il sera utilisé par spamassassin, à condition qu'on l'ait demandé dans le fichier de configuration de spamassassin.

Sur Linux-Mandrake, pyzor est disponible sous forme de paquetage directement installable (rpm). Son installation est donc automatisée.

4.2.4 DCC

DCC est une application du même type que razor. Si dcc est installé, il sera utilisé par spamassassin, à condition qu'on l'ait demandé dans le fichier de configuration de spamassassin.

Sur Linux-Mandrake, dcc est disponible sous forme de paquetage directement installable (rpm). Son installation est donc automatisée.

[2] Une même archive peu être utilisées autant de fois qu'on le souhaite pour "éduquer" spamassassin, sans risque de biaiser son jugement. En effet, spamassassin mémorise les courriels déjà utilisés. Quelques temps après le premier apprentissage, nos archives se seront enrichies de nombreux courriels. Nous pourrons donc relancer l'apprentissage sans crainte. Spamassassin reconnaîtra de lui-même quels sont les nouveaux messages.

4.2.5 IMAP Spam Begone

Le traitement par spamassassin des BAL consultées par IMAP requiert l'installation d'une application spécifique.

IMAP Spam Begone est un programme écrit en python. La plupart des distributions Linux installent python par défaut. Il suffit donc de récupérer le script isbg.py sur le site de IMAP Spam Begone (http://www.rogerbinns.com/isbg/),de le copier dans /usr/local/bin et de modifier ses propriétés pour le rendre exécutable. isbg n'a pas besoin d'être configuré car tous les paramètres sont fournis lors du lancement de l'application, en ligne de commande.

On va donc pouvoir tester immédiatement le fonctionnement d'isbg en tapant, dans un terminal ;

```
isbg.py --imaphost laposte.net --imapuser bibi.bobo --savepw
```

Le système vous demandera d'entrer votre mot de passe et le mémorisera.

Voici un exemple de lancement de l'application :

```
/usr/local/bin/isbg.py --imaphost imap.laposte.net --imapuser bibi.bobo --spaminbox spam -- ↩
    delete --expunge --verbose
```

On y indique le *nom du serveur IMAP*, le *nom de l'utilisateur* et le *nom du dossier distant* (spam) dans lequel `isbg` doit placer les messages suspectés d'être du pourriel. On ne précise pas le mot de passe puisqu'il a été mémorisé précédemment.

Au cas où vous utiliseriez plusieurs comptes IMAP, il est plus simple de créer un script qui lancera d'un seul coup toutes le connexions. On crée donc un fichier texte nommé imapspam que l'on place dans /usr/local/bin. Il contiendra autant de lignes que nous aurons de comptes IMAP à traiter et donc de lancements d'isbg à effectuer :

```
#!/bin/sh
/usr/local/bin/isbg.py --imaphost imap.laposte.net --imapuser bibi.bobo --spaminbox spam -- ↩
    delete --expunge --verbose
```

On prendra soin de rendre ce fichier exécutable. Par la suite, il suffira de taper `imapspam` pour lancer le tri de tous les comptes IMAP indiqués dans le script. Nous verrons, plus loin, que nous n'aurons même pas à taper cette commande car nous automatiserons la procédure.

4.3 Reporting

Cette opération consiste à informer les système coopératifs anti-spam des pourriels que nous avons reçus. En pratique, les applications nécessaires pour communiquer avec ces systèmes sont les mêmes que celles qui coopèrent avec spamassassin : razor, pyzor et DCC. Il n'y a donc rien de particulier à installer.

De plus, spamassassin effectue lui-même une partie du reporting. Les messages qu'il classe comme pourriel sont automatiquement signalés aux systèmes coopératifs. Seuls restent à signaler les pourriels qu'il n'a pas réussi à classer mais que l'utilisateur identifie comme pourriel. Grâce à ce signalement, le système de détection des spams sera plus efficace, à l'avenir.

4.4 Traque et plaintes

Le système de traque sert à débusquer les spammers et à les mettre hors d'état de nuire. J'ai retenu le système spamcop car il semble jouir d'une bonne réputation et l'inscription y est gratuite. Tout utilisateur souhaitant exploiter les services de spamcop doit préalablement s'enregistrer sur la page Web suivante : http://www.spamcop.net/mcgi?action=loginform

Le système spamcop propose deux outils complémentaires : un outil d'analyse de pourriels et un outil d'envoi de plaintes.

4.4.1 Analyse de pourriels

Cet outil va analyser le contenu d'un pourriel que vous lui soumettez et tenter de reconstituer les chemin suivi par ce message. Il repérera également les sites Web référencés dans le mél et, d'une manière générale, toute les adresses internet présentes. En exploitant sa connaissance de l'internet, spamcop vous proposer une collection d'adresses auxquelles envoyer une plainte concernant ce pourriel.

4.4.2 Envoi de plaintes

Envoyer une plainte avec sa propre adresse d'expéditeur, c'est s'exposer à des représailles de la part des spammers. C'est pourquoi spamcop vous propose d'envoyer lui-même vos plaintes en dissimulant votre identité. Attention, spamcop n'endosse pas la responsabilité des plaintes que vous émettez. Vous restez responsables de vos actes. D'ailleurs, spamcop sait parfaitement quel utilisateur émet quelle plainte et veille à éviter les plaintes calomnieuses ou injustifiées.

4.4.3 Fonctionnement

L'utilisation de spamcop se fait donc en deux temps.

Dans un premier temps, on envoie par mail, en pièce-jointe, les pourriels que l'on veut analyser. L'adresse à laquelle on expédie le courriel de traque est unique pour chaque utilisateur. Elle est donnée lors de l'inscription au service spamcop. En réponse, le service renvoie un courriel pointant vers des pages Web où l'on trouvera un bilan d'analyse pour chaque pourriels soumis.

Dans un second temps, on prend connaissance de ces résultat et l'on décide, ou non, d'envoyer une plainte aux différentes autorités plus ou moins impliquées par le pourriel analysé. Spamcop envoie alors un message d'avertissement aux autorités de que l'on aura retenues et leur fournit toutes les informations dont il dispose pour identifier nommément les auteurs du pourriel.

4.4.4 Limites d'utilisation

Comme on le constate, le lancement d'une traque n'est pas automatisée. Elle demande une implication réelle et volontaire de la personne qui lance la traque. C'est bien ainsi. Lancer une traque est une démarche comparable à un dépôt de plainte. Il est primordiale que le plaignant soit conscient de chacun de ses actes, de chacune de ses plaintes.

La lourdeur et les répercussions d'une telle procédure plaide pour qu'on l'utilise avec parcimonie. En pratique je ne l'utilise que sur des pourriels qui ont été ratés par les logiciels de tri. Si vous avez correctement réglé spamassassin et installé les programmes d'interrogation des services coopératifs, très peu de pourriels devraient passer au travers.

5 Installation et configuration

5.1 Fausse alerte (faux positif)

L'erreur la plus grave que peut commettre spamassassin est la fausse alerte. Dans ce cas, spamassassin a classer comme pourriel un message licite. L'erreur est fréquente, surtout au début, lorsque l'apprentissage est incomplet. Mais cette erreur peut également se produire avec un spamassassin aguerri. Il suffit que l'on reçoive une nouveau type de courriels licites qui auraient le mauvais goût de ressembler à des pourriels que l'on reçoit ou que l'on a reçu à un époque. Il va donc falloir entraîner spamassassin à discerner ces nouveaux messages.

L'apprentissage à partir de faux positifs soulève un problème technique. En effet, les messages classés par SA voient leurs entêtes modifiées. Si, comme moi, vous avez paramétré SA pour qu'il transforme les pourriels en pièces jointes, il va falloir rétablir ces messages dans leur forme original. En effet, on souhaite que l'apprentissage s'effectue sur un message totalement conforme au message tel qu'on la reçu, avant traitement.

Le dossier Faux-spam de Kmail contient donc des messages qu'il va falloir rétablir dans leur forme originale. Puis on pourra demander à SA d'apprendre à partir du contenu de dossier.

5.2 Liste noire, liste blanche

Spamassassin procède à une analyse très générale des courriels qui lui sont soumis. Il en résulte deux défauts majeurs : lenteur et fausse détection.

La lenteur n'est pas toujours justifiée. En effet, l'utilisateur sait parfaitement que les expéditeurs avec lesquels il communique régulièrement ne lui envoient jamais de pourriels. Il est donc inutile d'analyser leur messages. À l'inverse, certains expéditeurs ne lui envoie que du pourriel. Là encore, il est inutile de perdre du temps en analyse.

Le fonctionnement de spamassassin peut être accéléré si lui on fournit :

– une liste des expéditeurs dont on acceptera systématiquement les messages, ou liste blanche,
– une liste des expéditeurs dont on refusera systématiquement le courrier, ou liste noire.

Le système de liste blanche peut également s'avérer dans des situations où des expéditeurs identifiés vous envoient des messages valides que spamassassin s'obstine à classer comme pourriel[3]. Il s'agit d'une fausse détection. En plaçant ces utilisateurs connus en liste blanche, on contournera l'analyse défaillante de spamassassin.

6 Fonctionnement automatisé

6.1 Position du problème

Beaucoup d'utilisateurs de spamassassin (SA) trouvent le logiciel pratique, mais peu efficace. Leur jugement peut être résumé par la phrase suivante : "un peu de tri c'est mieux que rien, mais il y a beaucoup d'oublis et quelques erreurs, très regrettables". Sauf cas très exceptionnel, ce résultat médiocre est dû à une mauvaise configuration du logiciel. SA peut être incroyablement efficace mais il faut respecter deux conditions contraignantes

– il faut l'entraîner avec un volume considérable de courriels dont **beaucoup de ham**,
– il faut exploiter **tous** les modules additionnels.

La première condition ne peut être efficacement remplie que si l'on archive son courriel. La seconde peut en rebuter plus d'un. L'installation et la configuration de tous modules est une opération méticuleuse mais simple. C'est leur usage qui soulève de vraies difficultés. En effet, l'interrogation des serveurs peut être longue et peut rendre le logiciel de courriels presque inutilisable. Il s'agit, là encore, d'un défaut de configuration.

6.2 Un vrai système d'exploitation

La grande majorité des utilisateurs considère Linux comme une interface pour lancer des applications *interactives* de bureautique communicante : traitement de texte, courriel, navigation Web... Leur système leur apparaît comme des applications auxquelles on greffe des extensions. Il leur semble donc naturel d'installer spamassassin comme un plug-in de leur logiciel de courriel. C'est d'ailleurs ce que recommandent de nombreux modes d'emplois disponibles sur le Web.

Cette méthode est simple mais inefficace. Pourquoi faut-il absolument que la détection du pourriel se fasse au moment où l'on souhaite lire ou envoyer un courriel ? Or, c'est ce qui se passe. Si l'on reçoit beaucoup de pourriel, dès qu'on lance une collecte, notre logiciel de courrier reste bloqué pendant de longues minutes, indispensables à une détection efficace des pourriels. Bien sûr, on peut laisser notre logiciel "ouvert", en tache de fond, et le programmer pour qu'il relève le courrier, à intervalles réguliers. Mais est-ce vraiment ce que l'on souhaite faire ou bien est-ce le moyen détourné que nous avons trouvé pour ne pas être trop pénalisé ?

Avec Linux, on peut automatiser ce qui doit l'être et n'utiliser les logiciels interactifs que pour traiter des tâches interactives. Ce qu'on veut, c'est pouvoir ouvrir notre logiciel de courriel juste quand on en a besoin et que tous nos courriels soient déjà là, à portée de main, triés, catalogués. On veut que notre logiciel soit immédiatement disponible pour lire, rédiger, rechercher, classer nos courriels. Le filtrage anti-pourriels pourrait représenter un avantage pur, sans désagrément. Il le doit parce que c'est possible.

6.3 Passer la vitesse supérieure

Afin d'automatiser notre filtrage antispam nous allons ajouter deux applications principales. La première sera chargée d'effectuer la collecte du courrier, autrement dit, l'interrogation des serveurs pop ou IMAP par lesquelles nous avons accès au contenu de nos BAL.La seconde s'occupera du tri et de la classification du courrier, en faisant appel à spamassassin[4]

.On s'oriente ainsi vers une séparation des tâches qui va permettre de mieux réaliser chacune d'elles. On confiera l'ordonnancement des relèves périodiques à une troisième application, déjà présent dans le système (cron). Enfin, on devra utiliser un application spécialisée dans le transcodage de courriels, pour des raisons de compatibilité de formats.

[3] Cette erreur résulte généralement de courriels un peu tordus Ce peut être toujours le cas si votre correspondant utilise un mauvais logiciel de composition de courriels, l'horloge de son ordi n'est pas à l'heure, il n'utilise pas correctement la fonction citation, etc. Bref une foule de petit défauts, systématiques, mais totalement intégrés dans ses pratiques de travail.

[4] Notez bien qu'une fois ce mécanisme installé, on pourra l'étendre à d'autres fonctionnalités de tri des courriels, telles la détection et l'élimination des virus, par exemple.

Au final, notre logiciel de courriels ne nous servira plus qu'à consulter les courriels classifiés à composer des messages et à archiver nos courriels reçus ou expédiés. On néglige trop souvent cette dernière fonctionnalité, fondamentale, dès que le courriel et plus qu'un outil de "bavardage". C'est pour cette raison qu'on limitera le tri préalable des courriels au seul critère de la probabilité du pourriel. En effet, on aurait pu profiter de cette étape pour effectuer toutes les opérations de tri automatique que l'on fait traditionnellement réaliser par le logiciel de courrier : rangement des listes des dossiers séparés, classement des courriels par types (travail, amis, militantisme, etc.), classement par source (expéditeur ou BAL d'origine), etc. Tout logiciel de courriel (digne de ce nom) étant capable de remplir ces foncions vite et bien, on continuera à lui confier cette tâche, proche de l'archivage.

En pratique, le tri automatisé remplira trois boîtes aux lettres locales qui seront, ensuite relevées par notre logiciel de courriel. La première contiendra le ham, c'est-à-dire que ce l'on ne suspecte absolument pas pourvoir être du pourriel. Pour ce faire, nous utiliserons note BAL Unix (/var/spool/mail/bibi). La deuxième BAL locale contiendra les messages que spamassassin considère être des pourriels mais ne peut être sûrs à 110% qu'il en sont réellement. La troisième et dernière BAL locale contiendra les messages dont spamassassin nous affirme que ce ne peut raisonnablement par être autre chose des pourriels. Ces deux dernières BAL seront créés dans l'espace utilisateur (/home/bibi).

6.4 Application de collecte : fetchmail

L'installation de l'application de collecte est triviale sur une distribution Mandrake. On prendra soin d'installer conjointement `fetchmail` et son logiciel de configuration interactive `fetchmailconf`. Les deux sont disponibles sous forme de paquetages rpm.

La configuration sera simple, grâce à fetchmailconf. La plupart des paramètres à fournir sont évidents puisqu'ils reprennent les réglages classiques d'un logiciel de courriel. Pour chaque boîte aux lettres distante que l'on relève il faudra indiquer le serveur pop ou IMAP qui permet d'y accéder, le login d'utilisateur et le mot de passe qui en protègent l'accès..

Un paramètre, un peu particulier, vous permettra d'indiquer votre login d'utilisateur Linux. C'est une garantie contre la perte des courriels. En effet, si les opération ultérieures (de tri-classification) devaient générer des erreurs informatiques, il se pourrait que certains courriels soient perdus. Dans ce cas, fetchmail déposera les courriels qui n'ont pu être traités dans votre boîte aux lettre Unix, celle qui recueillira le ham et dont dispose tout utilisateur Linux, à savoir /var/spool/mail/<nom de login>.[5]

6.5 Application de tri : procmail

L'installation de l'application de tri est triviale, sur une distribution Mandrake. L'application `procmail` est disponible sous forme de paquetage rpm.

La configuration peut être très complexe si l'on veut exploiter à fond les possibilités de l'outil. Nous nous limiterons à lui faire activer spamassassin.

À la racine de notre espace de travail (/home/bibi), on créera un fichier de configuration dénommé `.procmailrc` et l'on y placera le contenu suivant :

```
SHELL=/bin/sh
MAILDIR=$HOME/MyMail
DEFAULT=$HOME/MyMail/inbox
LOGFILE=$HOME/.procmaillog
VERBOSE=yes

# On jette ce qui est déjà signalé comme du spam
# On suppose qu'il s'agit d'un spam certain...
:0
* ^Subject:.*\*\*\*SPAM\*\*\*
/dev/null

# On reformate les mails correctement, notamment le champ from initial
:0f
| formail
```

[5] La plupart des utilisateurs individuels d'ordinateurs Linux, ne se servent pas de cette BAL. Beaucoup n'en connaissent même pas l'existence puisqu'ils n'en ont pas besoin... Elle n'en existe pas moins.

```
# On soumet à l'analyse de spamassassin et prenant soin
# de nettoyer les spams analysés des précédentes traces d'analyse
:0fw
* < 256000
| spamassassin -d | spamassassin

# Si spamassassin se plante en cours de route, le mail
# qu'il traitait ne sera pas perdu. Il arrivera dans DEFAULT.
:0e
{
  EXITCODE=$?
}

# Le spam considéré comme 'certain' va dans la boîte qui lui est destinée
:0:
* ^X-Spam-Level: \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*
spam-certain

# Le spam considéré comme 'probable' va dans la boîte qui lui est destinée
:0:
* ^X-Spam-Status: Yes
spam-probable

# Les autres messages sont considérés comme licites et
# sont placés dans la BAL Unix
:0:
/var/spool/mail/$LOGNAME
```

Cette configuration lance l'activation de spamassassin, reformate correctement les courriels et les tri en fonction de leur probabilité d'être du spam.

NOTE
Sans entrer dans le détail, chaque ligne commençant pas ":0" indique le début d'une règle de "filtrage". Ainsi, la première règle dit simplement que qu'il faut reformater tout les courriels. La seconde réserve l'application de spamassassin aux messages d'une taille inférieure à 250 Ko[a]. La troisième règle, à usage informatique, permettra au système de savoir si quelque chose s'est mal passé pendant le traitement par spamassassin. Les régles suivantes effectuent le tri vers les BAL locales, en fonction des résultats du traitement de l'étape précédente. Le nombre de '*' figurant dans la quatrième règle fixe la *note* à partir de laquelle on décide de faire totalement confiance à spamassassin.

[a] 1 Ko vaut 1024 octet, et non 1000. Donc 250 Ko valent 256 000 octets et non 250 000.

6.6 Formater correctement les courriels : formail

Tout traitement automatique des courriels risque d'échouer si les messages qu'on lui soumet ne sont pas correctement formatés. On fera donc appel à l'application `formail` avant de procéder au moindre traitement. Sa fonction sera de faire en sorte que les courriers récupérés par fetchmail et les dossiers assemblés par procmail soient conformes et donc lisibles par tout logiciel de traitement de traitement automatique ou interactif de courrier, tel Kmail.

L'installation de l'application est triviale sur une distribution Mandrake. L'application `formail` est disponible sous forme de paquetage rpm.

Cette application ne requiert aucune configuration. Comprenez : la configuration de base, préinstallée, nous conviendra. Il aura suffit de la mentionner dans le fichier `.procmailrc` pour qu'elle soit utilisé à bon escient.

6.7 L'ordonnanceur : cron, anacron

Lancer des travaux à heures régulières ou à intervalles régulier est une spécialité. Plutôt que d'obliger chaque logiciel qui a besoin d'un tel mécanisme à le prévoir, il est plus simple que tous utilisent le même "ordonnanceur". Cet outils, `cron`, est déjà installé

dans votre système et probablement déjà utilisé par des applications, sans que vous le sachiez.[6]

Mais `cron` a été conçu pour un ordi que l'on n'arrête jamais, tel un serveur fonctionnant 24h/24 et 7j/7. Or, la plupart des ordis individuels sont éteints une grande partie du temps. On installera donc un autre logiciel, mieux adapté à notre utilisation sporadique : `anacron`. Cette application existe sous forme de paquetage rpm et ne nécessite aucune configuration particulière. Une fois installée, on l'oublie complètement et on configure l'application `cron`, comme si anacron n'existait pas[7]. `Anacron` saura, toute seule, aller récupérer les paramétrages de `cron`, lancera les opérations programmées qui n'ont pu être effectuées pour cause d'extinction et se chargera de tout.

L'installation de l'application est triviale sur une distribution Mandrake. L'application `anacron` est disponible sous forme de paquetage rpm. Aucune configuration n'est nécessaire. On se contentera de paramétrer cron.

On accède paramètres de cron par la commande `contrab -e`. Elle provoque l'ouverture de votre éditeur de texte par défaut sur le bon fichier. Il ne vous reste plus qu'à indiquer les tâches que vous souhaitez planifier :

```
*/30 * * * * /usr/local/bin/imapspam >> .imapspam.log
*/15 * * * * /bin/date >> .fetchmail.log ; /usr/bin/fetchmail -m /usr/bin/procmail >> . ↩
    fetchmail.log
5 0 * * * /usr/sbin/logrotate -s $HOME/.logrotate.status $HOME/.logrotate.conf
```

Chaque ligne correspond à une tâche. Ici, la principale tâche est la deuxième. Elle commence par insérer la date dans le journal des action (log) puis elle indique à cron qu'il doit, toutes les 15 minutes, lancer la collecte du courrier et faire traiter par procmail les courriers ainsi collectés . Les comptes-rendus de collecte seront placés dans le fichier `.fetcmail.log`.

La troisième tâche se charge précisément de traiter les fichiers de comptes-rendus (log). En effet, ces fichiers ne font que grossir. Si l'on n'y prend garde, il finiront par dévorer notre espace libre. On fait donc appel à une application spécialisée, `logrotate`, qui se charge de limiter intelligemment l'espace utilisé. Cette tâche est lancée, tous les jours, à 0h05. Si, à cette heure là, l'ordi est éteint, anacron lancera cette tâche lorsqu'on redémarrera l'ordi.

La première tâche, lance le tri à distance des boîtes que l'on consulte par IMAP. Toutes les 30 minutes, on active le script `imapspam`, vu précédemment. Bien sût cette ligne n'est nécessaire que si l'on a besoin d'imapspam !

6.8 Un peu de nettoyage : logrotate

Les fichiers de comptes-rendus (log) sont bien utils mais finissent par devenir encombrants. Afin de limiter l'espace consommé, nous avons décider de limiter la taille de tous les fichiers de log que nous avons créés. Pour ce faire, nous faisons appel à une application spécialisée dans ce type de travail, `logrotate`.

L'installation de l'application est triviale sur une distribution Mandrake. L'application logrotate est disponible sous forme de paquetage rpm. On la configure en créant un fichier de paramètres. Le choix du nom est laissé à la discrétion de l'utilisateur, il faut simplement indiquer ce nom lorsqu'on lance la command logrotate. Comme nous avons confier son lancement à cron, nous appellerons ce fichier conformément à ce que nous avons indiqué dans le fichier de commande de cron : `/home/bibi/.logrotate.conf`. Pour chaque fichier de log, nous indiquerons quelle politique de limitation de l'espace nous voulons voir appliquer :

```
$HOME/.fetchmail.log {
 rotate 1
 size=50k
}
$HOME/.procmaillog {
  rotate 1
 size=100k
}
$HOME/.imapspam.log {
  rotate 1
 size=50k
}
```

[6] Sur Mandrake, si vous avez demandé que le système effectue une sauvegarde hebdomadaire de vos données, c'est `cron` qui lancera l'opération de sauvegarde, le moment voulu.

[7] On ne tripatouille le fichier /etc/crontab avec son éditeur préféré. Pour entrer les ordres dans crontab, on utilise la commande "`crontab -e` ".

Ici, nous adaptons la même politique pour les trois fichiers de log que nous traitons. Nous indiquons que leur taille ne doit pas dépasser 50Ko ou 100Ko, selon le fichier concerné. Dès qu'elle dépasse la limite fixée, logrotate effectue une copie de sauvegarde du fichier et le remplace par un fichier vierge. "rotate = 1" signifie qu'on ne fera pas de sauvegarde des sauvegardes (1 niveau de sauvegarde), chaque sauvegarde venant écraser la précédente. Donc, pour chaque fichier géré ici, dans le pire des cas, l'espace maximum consommé sera égal à deux fois la limite fixée : un fois pour le fichier et une fois pour la sauvegarde.

7 Fiche de travail

J'ai résumé ici la liste des opérations à effectuer, ainsi que leur description pratique. L'idée est qu'un utilisateur qui exploiterait le même environnement que moi, pourrait configurer son système en suivant, pas à pas, l'ensemble des indications fournies. Les utilisateurs d'autres environnements devront interpréter et adapter ces indications.

Les boîtes consultés par IMAP sont un cas à part. Si on les **relève**, comme des BAL pop, on les traitera avec fechmail (comme les BAL pop). Si on les **consulte**, sans les relever, on devra continuer à les trier à distance, avec `isbg`. Cron nous permettra simplement d'automatiser ce tri, si bien que nos BAL IMAP nous sembleront toujours triées, lorsque nous les consulterons via Kmail.

7.1 Installation des logiciels

Nous allons commencer par installer tous les logiciels dont nous aurons besoin. Certains sont, peut-être, déjà installés sur votre ordi mais peu importe. Pour faire court, je vais supposer que l'on effectue les installations en ligne de commande. Si vous avez l'habitude d'utiliser l'interface graphique de Mandrake, vous n'aurez aucun mal à transposer. A chaque fois, accepter d'installer toute les dépendances proposées

A l'exception d'un seul logiciel (léger) qui demande un accès à internet, vous devriez pouvoir tout installer à partir des 4 CD de la Mandrake 10.0 Communitiy. Bien entendu, vous aurez inévitablement à faire des mises à jour, après l'installation. Donc, si vous le pouvez, je vous conseille de faire l'installation à partir de sources Mandrake disponibles en ligne (update et contrib). Si vous disposez d'un accès rapide, je vous conseille même de désactiver vos sources CD.[8] et de les remplacer par une source FTP. Ceci vous dispensera de toute manipulation de CD[9].

Allons-y...

1. Ouvrir une fenêtre de terminal
2. Se connecter en tant que root (la commande demandera le mot de passe de root) :
   ```
   > su
   ```
3. Installer les outils de base
   ```
   > urpmi kate
   > urpmi anacron
   > urpmi python
   > urpmi logrotate
   ```
4. Installer spamassassin et ses auxiliaires
   ```
   > urpmi spamassassin
   > urpmi razor
   > urpmi pyzor
   > urpmi dcc
   ```
5. Installer les logiciels de traitement du courrier électronique
   ```
   > urpmi kmail
   > urpmi fetchmail
   > urpmi fetchmailconf
   > urpmi procmail
   ```

[8] Ne les supprimez pas, elles pourraient s'avérer vitale en cas problème... d'accès internet, par exemple ;-)

[9] Si tout cela vous paraît trop compliqué, faites avec les CD, comme vous en avez l'habitude, ça marchera très bien.

6. Installer le serveur pop et IMAP
   `> urpmi imap` (pour Ubuntu, installer dovecot)
7. Installer IMAP Spam BeGone (uniquement si vous consultez des BAL en IMAP)
   ouvrir l'url http://www.rogerbinns.com/isbg/
   y récupérer le fichier isbg.py et se placer dans le répertoire où on l'a téléchargé
   ```
   > cd le_répertoire_en_question
   > cp isbg.py /usr/local/bin
   > chmod a+x /usr/local/bin/isbg.py
   ```

Ça y est, vous avez tout installé.

7.2 Configuration de base

À partir d'ici, je supposerai que votre accès à internet est opérationnel et que kmail (ou équivalent) a été configuré et fonctionne.

7.2.1 Spamassassin

1. Se déconnecter en tant que root
   `> exit`
2. ouvrir l'url : http://www.yrex.com/spam/spamconfig.php
3. Remplir les champs (les valeurs par défaut conviennent mais faites au plus près de vos besoins)
4. Générer le fichier de configuration
5. Télécharger le fichier de configuration et l'enregistrer dans `$HOME/.spamassassin/user_prefs` ($HOME désigne votre répertoire de base.

7.2.2 Razor

1. Créer les fichiers de configuration
   `> razor-admin -d -create`
2. S'enregistrer auprès de razor (opération entièrement automatique)
   `> razor-admin -register`

7.2.3 Fetchmail

Attention ! Vous allez déclarer quelles BAL vont faire l'objet d'une relève et d'un filtrage automatique. Je vous conseille de faire comme moi et de n'enregistrer qu'une seule BAL, correspondant à une adresse anonyme qui ne vous sert à rien. Si vous n'avez pas une telle adresse électronique, je vous conseille vivement d'en créer une[10] *et de n'utiliser que celle-ci pendant les tests de votre installation.*

1. Lancer l'utilitaire de configuration
   `> fetchmailconf`
2. Comme sur un logiciel de courrier classique, créer les comptes correspondant aux BAL que l'on veut relever

7.2.4 Procmail

1. Ouvrir le fichier qui contiendra les paramètre de procmail
   `> kate $HOME/.procmailrc`
2. Copier-coller l'exemple de fichier .procmailrc présenté au chapitre précédent

[10] Le site de La Poste permet de créer instantanément une BAL. consultable en POP.

7.2.5 Cron

Attention ! À partir du moment où vous aurez configuré cron, la relève automatique de votre courrier commencera.

1. Ouvrir son fichier personnel de tâches pour cron
   ```
   > export EDITOR=kate
   > crontab -e
   ```
2. Copier-Coller l'exemple de crontab donné au chapitre précédent

7.2.6 logrotate

1. Ouvrir le fichier personnel de configuration de logrotate
   ```
   > kate $HOME/.logrotate.conf
   ```
2. Copie-Coller l'exemple de fichier donné au chapitre précédent

7.2.7 Kmail

Je suppose que Kmail fonctionne, en utilisation classique. Il s'agit donc d'en adapter la configuration à la relève des trois boîtes locales que nous utilisons.

Dans Kmail, la création d'un compte associé à une BAL locale est triviale. Lorsqu'on crée un compte, il suffit de sélectionner le bouton "*boîte aux lettres locale*". Il faudra ensuite indiquer quel fichier (local) joue ce rôle de boîte aux lettres et préciser quel mode de protection on retient. Puisqu'on utilise le logiciel `procmail` on optera pour la méthode intitulée "*fichiers de verrouillage procmail*".

1. Créer un dossier dénommé SPAM
2. Dans ce dossier créer deux boîtes de type "mbox" que l'on appellera : `X_Certain` et `X_Probable`.
3. Créer le compte de boîte locale associé au fichier `$HOME/MyMail/Spam-certain` et demander à ce que son contenu arrive dans X_Certain.
4. Créer le compte de boîte locale associé au fichier `$HOME/MyMail/Spam-probable` et demander à ce que son contenu arrive dans X_Probable.
5. Créer le compte de boîte locale associé au fichier `/var/spool/mail/<user_name>` et demander à ce que son contenu arrive dans la Boîte de réception.

Pour chacun de ces comptes on définira une périodicité de relève automatique de 17[11] minutes.

7.3 Améliorations

L'efficacité de SA repose en grande partie sur la qualité de son apprentissage. Pour que le travail d'éducation ne devienne pas tellement lourd qu'on en vienne à le négliger, il est primordial de le simplifier.

7.3.1 Déclaration immédiate des faux positifs

L'erreur la plus grave que peut commettre SA est le faux-positif. On s'impose donc de les lui signaler tous.

1. Dans Kmail, dans le dossier SPAM créé précédemment, on crée une boîte de type *mbox*, dénommée `Faux-spam`.
2. Se connecter en tant que root
   ```
   > su
   ```
3. Ouvrir un fichier de texte et y déposer le texte ci-joint
   ```
   > kate
   ```

[11] Pourquoi 17 ? Parce que c'est le premier nombre premier après 15, qui est la périodicité de relève des boîtes distantes.

```
#!/bin/sh
SPAMDIR=$HOME/.Mail/.SPAM.directory/
cat $SPAMDIR'Faux-spam' | formail -s spamassassin -d >> ~/.faux-spam.tmp
rm -f $SPAMDIR'Faux-spam'
rm -f $SPAMDIR'.Faux-spam.index'
rm -f $SPAMDIR'.Faux-spam.index.ids'
rm -f $SPAMDIR'.Faux-spam.index.sorted'
mv ~/.faux-spam.tmp $SPAMDIR'Faux-spam'
chmod 600 $SPAMDIR'Faux-spam'
sa-learn  --ham --mbox $HOME/.Mail/.SPAM.directory/Faux-spam
```

4. Enregistrer le fichier dans /usr/local/bin/fauxspam
5. Le rendre exécutable
   `> chmod a+x /usr/local/bin/fauxspam`

Pour utiliser cette facilité :

1. dans Kmail, on déplacera les courriels s'avérant être des faux positifs dans le dossier éponyme.
2. Pour plus de sécurité, on quittera Kmail
3. on lancera la commande fauxspam

Cela aura un double effet. D'un part, spamassassin ne commettra plus le même genre d'erreur. Mais, de plus, le faux-positif aura retrouvé l'allure normale pour un courrier licite abandonnant la forme peu exploitable[12] que donne SA aux messages qu'il assimile à des pourriels.

7.3.2 Le faux négatif

Beaucoup moins grave que le faux positif qui risque ne nous faire manque un vrai courriel, le faux négatif est tout de même perturbateur. On a donc intérêt à les déclarer à SA, pour qu'il apprenne de ses erreurs. On en profitera pour lui confirmer la validité de ses choix concernant les vrai spams et, tant qu'on y est, on informera les systèmes coopératifs de tout cela.

1. On profite du fait qu'on est encore connecté en tant que root pour créer un nouveau fichier de commande
   `> kate /usr/local/bin/spamreport`
2. copier coller le contenu ci-joint

```
#!/bin/sh
SPAMDIR=/$HOME/.Mail/.SPAM.directory/
#Quit
# Reporting des pourriels correctement triés et *vérifiés* !
#
echo -e "Reporting du Spam Certain\n"
cat $SPAMDIR'X_Certain' | formail -s spamassassin -d | razor-report
#
#Reporting et apprentissage des pourriels oubliés
#
echo -e "Apprentissage et reporting du Spam Oublié\n"
cat $SPAMDIR'Oublié' | formail -s spamassassin -r
echo -e "Reporting et apprentissage Terminés\n"
#
# Vidange des boîtes-aux-lettres
#
rm -f $SPAMDIR'X_Certain'
rm -f $SPAMDIR'.X_Certain.index'
rm -f $SPAMDIR'.X_Certain.index.ids'
rm -f $SPAMDIR'.X_Certain.index.sorted'
touch $SPAMDIR'X_Certain'
chmod 600 $SPAMDIR'X_Certain'
rm -f $SPAMDIR'Oublié'
rm -f $SPAMDIR'.Oublié.index'
```

[12] Précisément pour limiter leur exploitation, pour des raisons de sécurité.

```
rm -f $SPAMDIR'.Oublié.index.ids'
rm -f $SPAMDIR'.Oublié.index.sorted'
touch $SPAMDIR'Oublié'
chmod 600 $SPAMDIR'Oublié'
echo -e "Boîtes vidées\n"
```

3. Enregistre le fichier, puis le rendre exécutable
   `> chmod a+x /usr/local/bin/spamreport`
4. Se déconnecter en tant que root
   `> exit`
5. dans Kmail, créer une boîte mbox nommée `Oublié` dans le dossier `SPAM`
6. quitter Kmail

Pour utiliser ces nouvelles facilités :

1. Dans Kmail, on déplacera les pourriels non détectés par SA dans le dossier Oublié
2. On décidera du sort réservé au spam probable. Soit on le déplacera dans le spam certain, après avoir vérifier qu'il ne contient vraiment que du spam. Soit on le videra, après avoir vérifié qu'il ne contient aucun faux positif. Soit ou vidangera le dossier sans y faire attention.
3. Pour plus sécurité, on quittera Kmail
4. Si l'on est sûr que les dossiers de Kmail `X_certain` et `Oublié` contiennent exactement ce qu'ils sont sensés contenir[13], on lance, dans un terminal, la commande `spamreport`.

7.4 Ouverture d'un compte SpamCop

Spamcop.net propose divers services liés à la lutte anti-spam. La plupart sont payants mais il existe un service auquel l'inscription reste gratuite : http://www.spamcop.net/anonsignup.shtml.

Une fois inscrit(e) vous disposerez d'un identifiant et ils vous donneront une adresse mél qui vous permettra de leur envoyer; par courriel, les messages que vous voulez analyser.

Le principe est le suivant :

1. Vous recevez un pourriel pour lequel vous souhaitez "porter plainte"
2. Réexpédiez-le, tel quel, en pièce jointe d'un message vide, à l'adresse mél qui vous a été attribuée
3. Vous recevrez, par mél, une notification d'analyse du message soumis. Cette notification contient un url qui vous permet d'accéder à la page Web contenant les résultat d'analyse
4. On vous propose d'envoyer un message d'alerte aux différents intermédiaires qui ont été impliqués dans l'acheminement du pourriel.
5. Vous sélectionnez décochez ce que vous voulez et vous validez l'envoi de l'alerte. Vous n'avez rien à rédiger.

Si la procédure paraît lourde, en fait, à chaque étape, on ne fait de donner quelques clics de souris.

7.5 Récapitulons

Si l'on fait le bilan de ce que nous avons vu et installé :

– Les boîtes que nous consultons en IMAP sont automatiquement triées à distance et le spam est déplacé du dossier `INBOX` au dossier `spam`

– Les boîtes que nous relevons en POP ou en IMAP sont automatiquement relevées et leurs contenus sont fusionnés et ventilés dans trois boîtes locales. Ces boîtes qui correspondent à des niveaux de présomption de pourriel, sont automatiquement relevées par Kmail.

– Les commandes manuelles **spamreport** et **fauxspam** nous permettent d'éduquer SA et de participer à la lutte collective contre le spam.

[13] X-certain du pourriel avéré et Oublié des faux négatifs

8 Consultation distante

Grâce au rapatriement de tous nos courriels, depuis nos différents comptes, sur notre poste de travail, nous avons pu les trier correctement. Cela suppose que nous consultions toujours nos courriels depuis notre poste de travail. Or, lorsque nous nous déplaçons (rien qu'en allant au travail), nous voudrions pouvoir continuer à consulter nos courriels.

Puisque notre ordi sait relever toutes nos BAL et classer les pourriels, l'idéal serait qu'on ait la possibilité de consulter, à distance, la BAL des courriels licites. Ainsi, même si on dispose d'une accès à bas débit, on aura un accès rapide à son "bon" courrier et on ne sera pas gêné par le flot des pourriels.

8.1 Installer un serveur POP/IMAP

Aussi incroyable de que cela puisse paraître, il est trivial de transformer son ordi sous GNU-Linux en serveur POP/IMAP. Il suffit d'installer l'application `imap`, disponible sous forme de paquetage rpm. Après l'installation automatique, aucune configuration n'est nécessaire. Oui, vous avez bien lu !

NOTE
Sous Ubuntu (base debian) l'application `dovecot` fera la même chose.

Tout utilisateur référencé sur cet ordi Linux pourra accéder, depuis n'importe quel ordinateur raccordé à l'internet, au contenu de BAL Unix[14]. Pour y accéder, le login et le mot de passe seront ceux que vous devez taper pour ouvrir une session Linux. C'est aussi simple que cela. Il n'y a rien à configurer car la configuration de base d'imap (ou dovecot) est celle que nous voulons.

C'est bien gentil tout ça, mais que devrais-je indiquer comme nom de serveur pop ou imap, lorsque je voudrai consulter mes mails depuis l'extérieur ? C'est là que les choses se compliquent plus ou moins, selon la manière dont vous raccordée à un internet. Avant de pouvoir s'identifier (et relever son courrier) il faut déjà pouvoir contacter le serveur.

8.2 Contacter son ordi, devenu serveur

8.2.1 IP fixe

Si votre formule d'accès à internet vous alloue une IP fixe, vous pourrez taper ce numéro d'IP à la place du nom du serveur. Là où vous avez l'habitude d'écrire pop.fournisseur.fr, vous taperez simplement 87.128.65.78, à supposer que telle soit l'IP qui vous est allouée. Vous donnerez le même numéro IP pour accéder en POP ou en IMAP.

Si l'utilisation d'une adresse numérique vous rebute, vous pouvez vous inspirer des recommandations données pour le traitement des IP dynamiques. Vous y verrez comment on peut associer un nom à l'IP que vous a allouée votre FAI.

8.2.2 IP dynamique

Mais la plupart des offres grand public d'accès à internet n'incluent pas une IP fixe. On parle alors d'IP dynamique, ce qui signifie que votre fournisseur d'accès vous allouera un numéro IP différent à chaque connexion. Même si vous ne vous déconnectez pas volontairement, votre FAI vous déconnectera probablement, au bout de quelques heures. Votre IP changera donc inévitablement, que votre accès permanent se fasse par modem téléphonique classique par réseau câblé ou par ADSL.

Il y a donc deux problèmes à résoudre. Le premier et de rétablir automatiquement la connexion, en cas de coupure. Le second est de trouver un moyen d'identifier votre ordinateur de manière permanente, malgré son changement périodique d'IP. Bien que ceci dépasse le cadre de cet article, j'en dirai quand même deux mots.

La continuité de connexion ne peut être assurée, malgré les déconnexions, que si vous disposez d'un mécanisme (logiciel) de détection des déconnexion et de reconnexion automatique. Dans le cas d'un accès ADSL, cette double fonctionnalité peut être directement intégré au modem/routeur ou au routeur par lequel vous êtes relié à votre fournisseur d'accès. Dans le cas le plus général, ce sera une fonctionnalité de l'utilitaire d'établissement de connexion, disponible sur votre ordi.

[14] Celle qui se trouve dans /var/spool/mail/<login utilisateur>.

Une fois l'ordi connecté en permanence, reste à l'identifier sur l'internet, depuis n'importe quel ordi du réseau des réseaux. Le moyen le plus simple est de faire appel à un service de gestion des adresses dynamiques, tel dyndns.org. Il vous permettra de déclarer un nom de domaine tel que `domainebibi.dyndns.org`. Pour que le service sache quel numéro IP associer à ce nom, il faudra inévitablement que vous installiez, sur votre ordi, un utilitaire[15] qui avertira le service (ici, dyndns.org) de tout changement d'IP détecté par l'ordi.

Devant la demande croissante pour ce type de service d'adresse dynamique, de plus en plus d'utilitaires de connexion intègre cette fonctionnalité, conjointement à la reconnexion automatique. De même, un nombre croissant de modem/routeur et de routeur intégrent cette fonctionnalité

Pour vous connecter, à distance, au serveur POP/IMAP installé sur votre ordi, il suffira alors d'indiquer le nom de domaine que vous avez déclaré auprès de votre service de gestion d'adresses dynamiques.

8.3 Franchir les protections

L'utilisation d'un accès permanent à internet vous expose à l'agression potentielle permanente. Si vous êtes un minimum prudent, vous aurez probablement protégé votre ordi derrière un firewall. Ce dispositif ayant pour but d'empêcher les accès non désirés à votre ordi, il y a de fortes chances pour qu'il bloque toute tentative de connexion IMAP ou POP provenant de l'internet[16].

Si vous souhaitez accéder à distance à votre BAL Unix se trouvant sur votre ordi, vous devrez donc explicitement paramétrer votre firewall pour qu'il laisse passer vos requêtes. En pratique cela reviendra à ouvrir les ports TCP 110 et TCP 143 qui correspondent, respectivement, aux services POP et IMAP.

Notez bien que cela signifie que vous ouvrez une brèche dans votre protection. Ceci n'est pas catastrophique. Utiliser un ordinateur conduit immanquablement à mettre en danger sa sécurité... Le seul ordi 100% sur, est celui qu'on n'allume jamais, et encore ; il peut disparaître dans un incendie ou être emporté par une inondation ;-) Soyez sérieux. Ouvrir volontairement une brèche ne signifie pas baisser la garde. Dès lors, sachez que la sécurité de votre système repose sur sa solidité et sur la fiabilité du logiciel gérant le service POP/IMAP. Heureusement, en choisissant GNU-Linux, vous avez opté pour un système présentant de bonnes garanties de sécurité[17].

Pour plus de sécurité, vous préférerez utiliser IMAPS à IMAP. Vous n'aurez qu'à changer le réglage du logiciel grâce auquel vous effectuez la consulation à distance du courrier collecté par votre ordi. `imap` et `dovecot` sont déjà configurés pour fonctionner avec IMAPS.

9 Questions-réponses

Pourquoi ai-je utilisé procmail pour lancer spamassassin et non postfix ou sendmail ?

Lancer le contrôle des pourriels depuis le logiciel d'acheminement des courriels (MTA - Mail Transport Agent) n'a de sens que si l'on opère un filtrage de base, valable pour tous les utilisateurs d'un système informatique. Dans notre cas, nous voulons appliquer des filtres aussi fins que possibles, donc parfaitement adaptés à un utilisateur particulier. Cela implique que soit utilisée la base d'apprentissage personnelle d'un utilisateur. Le but n'est donc pas tant d'invoquer SA depuis procmail que de l'invoquer en tant qu'utilisateur individuel. L'appel de procmail se faisant depuis la crontab de l'utilisateur, il se fera sous son identité. procmail transmettra cette identité à tous les programmes qu'il appellera, notamment spamassassin. SA utilisera donc les fichiers de configuration de cet utilisateur particulier.

Pourquoi avoir placé certains courriels dans la BAL Unix et d'autres dans l'espace utilisateur ?

C'est uniquement pour simplifier la configuration de la consultation distante. Ainsi, l'installation du logiciel gérant l'accès POP et IMAP (`imap` ou `dovecot`) ne requiert aucune configuration. Elle donne naturellement accès à la BAL Unix. Si un utilisateur, n'est pas intéressé par la consultation à distance et s'il souhaite regrouper tous ses messages dans son espace personnel, il lui suffit d'adapter une ligne de son fichier .procmailrc et un compte Kmail de relève des boîtes locales.

Pourquoi ne pas confier à cron de vider régulièrement le dossier spam-certains ?

[15] ddclient, disponible sous forme de rpm pour Mandrake (ou de paquettage pour Ubuntu) remplit parfaitement cette fonction. Elle est installée avec un fichier de configuration qu'il faut adapter au service que vous avez choisi (dyndns ou autre).

[16] Les connexion POP ou IMAP partant de votre ordi à destination de l'internet ne sont pas bloquées. Heureusement, sinon vous ne pourriez pas relever vos BAL depuis votre ordi...

[17] Bien sûr, il y a mieux (openBSD, par exemple), mais il y surtout bien pire (Window$, toutes versions confondues, évidemment).

Me plaçant dans une optique de lutte (pas simplement de protection), je préfère déclarer les messages que je considère comme étant des pourriels plutôt que de les jeter à la poubelle. C'est une contribution minimale à l'effort collectif de lutte anti-pourriels. S'agissant d'une déclaration dont le conséquences peuvent être graves pour les émetteurs, j'estime qu'une telle action ne peut être confiée à un programme. Un utilisateur doit toujours être pleinement conscient de ce qu'il déclare quand il affirme "ceci est un pourriel et son émetteur est un spammer !"

10 Liens utiles

Le site principal de spamassassin : http://spamassassin.org/

Le wiki qui renferme une documentation très complète expliquant le fonctionnement de spamassassin ainsi que les réponses aux questions les plus courantes (les simples et les compliquées) : http://wiki.apache.org/spamassassin/

Une description claire et directe de la configuration de spamassassin et des trois modules externes (razor, pyzor et dcc), en vue de leur coopération : http://www.konabi.de/content.php?action=spamassassin