

Filtrage des courriers non désirés

**Alain Patrick AINA
(aalain@trstech.net)**

Abidjan, le 06 Avril 2006

Sources des courriers non désirés

- Spam
 - Courriers commerciaux envoyés en bloc sans consentement des destinataires
 - Souvent frauduleux- i.e. “élargissement de pénis”, “Les offres de loterie”, offres de produits pharmaceutiques, relatif aux anciens presidents africains....
- Virus, Les “cheval de troie”
 - Machine infectée envoyant des courriers sans le consentement du propriétaire
- Les rebonds malicieux ("Joe-jobs")
 - Spam ou virus envoyés avec “MAIL FROM” forgé

Les rebonds vont à la tierce partie innocente
- Déluge de réponses à des mails forgés

Quels en sont les coûts?

- Des courriers importants accidentellement rejetés dans une forêt de pourriels
- Perte de temps
 - Suppression des pourriels
 - Configuration et maintenance des filtres
 - Consultation des courriers rejetés pour y détecter de bons
- Perte de bande passante et d'espace disque
 - Spécialement pour les utilisateurs de modems
 - Les pièces jointes des virus et spams peuvent être grands
- Désagrément, offenses, et même fraude

Simple Mail Transfer Protocol

(SMTP)

Message en transit

- Un message est transmis par une *enveloppe*
MAIL FROM:< <didier.kla@citelecom.ci>
RCPT TO:<aalain@trstech.net>
- L'enveloppe est séparée du message RFC 2822
- Les champs de l'enveloppe (RFC 2821) n'ont pas besoin d'être identiques aux champs de l'en-tête (RFC 2822)
- Les MTA sont concernés par les enveloppes
Juste comme la poste...
- Les messages d'erreur ont un champ expéditeur nul
MAIL FROM:<>

Envelope expéditeur et entêtes(1)

```
220 alain.trstech.net ESMTP Exim 4.60 Tue, 11 Apr 2006 11:23:20
EHLO alain.trstech.net
250-alain.trstech.net Hello alain.trstech.net [62.56.186.219]
250-SIZE 52428800
250-PIPELINING
250 HELP
MAIL FROM:<aalain@trstech.net> SIZE=471
RCPT TO:<aalain@trstech.net>
DATA
250 OK
250 Accepted
354 Enter message, ending with "." on a line by itself
From: "AINA ALAIN PATRICK(TRS)" <aalain@trstech.net>
Reply-To: aalain@trstech.net
Organization: www.trstech.net
To: aalain@trstech.net
Date: Tue, 11 Apr 2006 11:22:12 +0000
User-Agent: KMail/1.8
MIME-Version: 1.0
Content-Type: text/plain;
  charset="us-ascii"
Content-Transfer-Encoding: 7bit
Content-Disposition: inline
Message-Id: <200604111122.12642.aalain@trstech.net>
```

--

See you at AfNOG 2006/AfriNIC-IV
Nairobi, 7-17 may 2006

Envelope expéditeur et entêtes(2)

Return-path: <parvate@kvalvaag.net>

Envelope-to: aalain@trstech.net

Delivery-date: Tue, 04 Apr 2006 23:26:45 +0000

Received: from [200.87.232.15] (helo=kvalvaag.net)

by macl.netcom.tg with smtp (Exim 4.34)

id 1FQuuv-0007AF-Hi

for aalain@trstech.net; Tue, 04 Apr 2006 23:26:45 +0000

Message-ID: <000001c6583a\$7493f980\$e061a8c0@hda80>

Reply-To: "Parvati Moorhead" <parvate@kvalvaag.net>

From: "Parvati Moorhead" <parvate@kvalvaag.net>

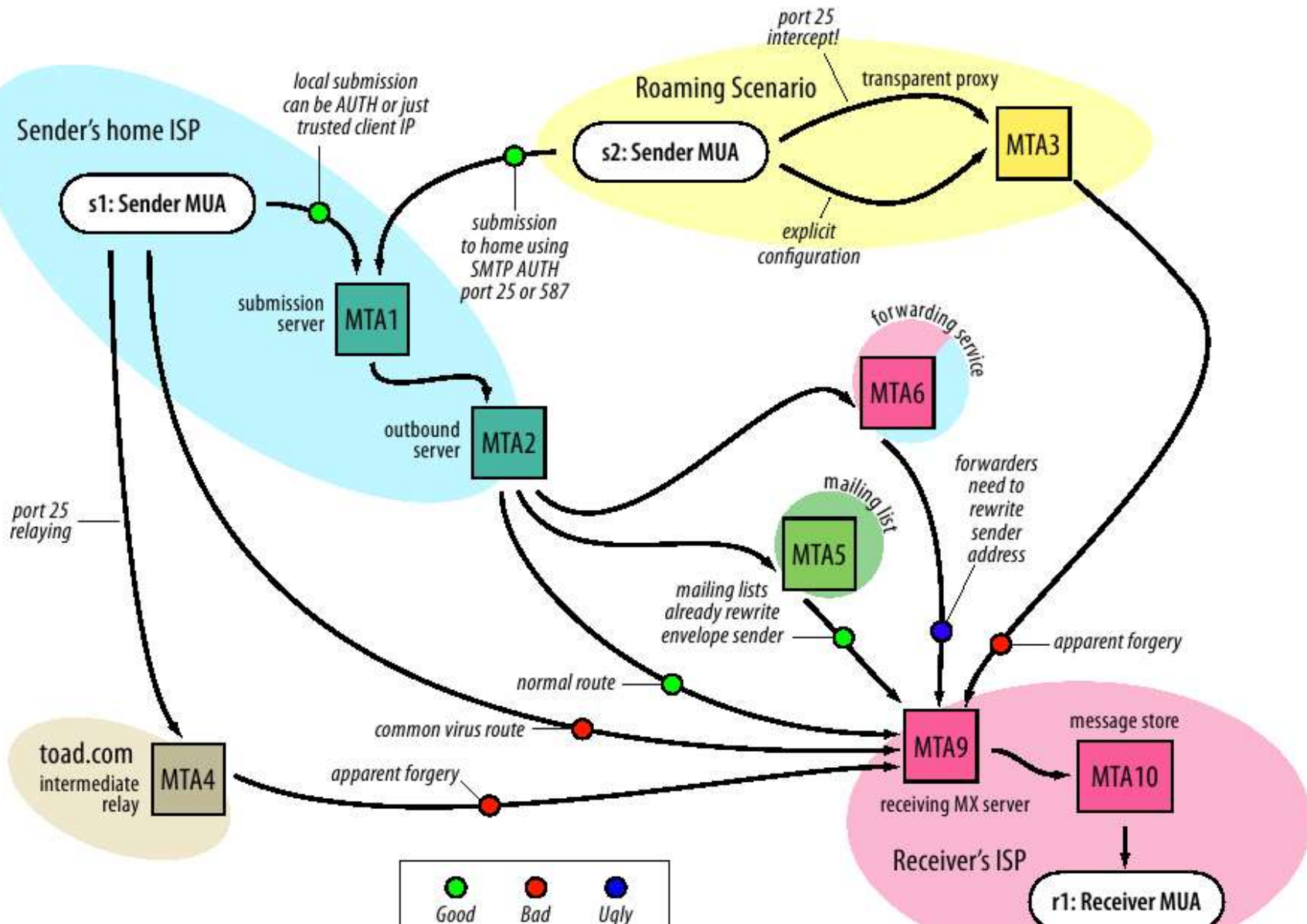
To: aalain@trstech.net

Subject: look new

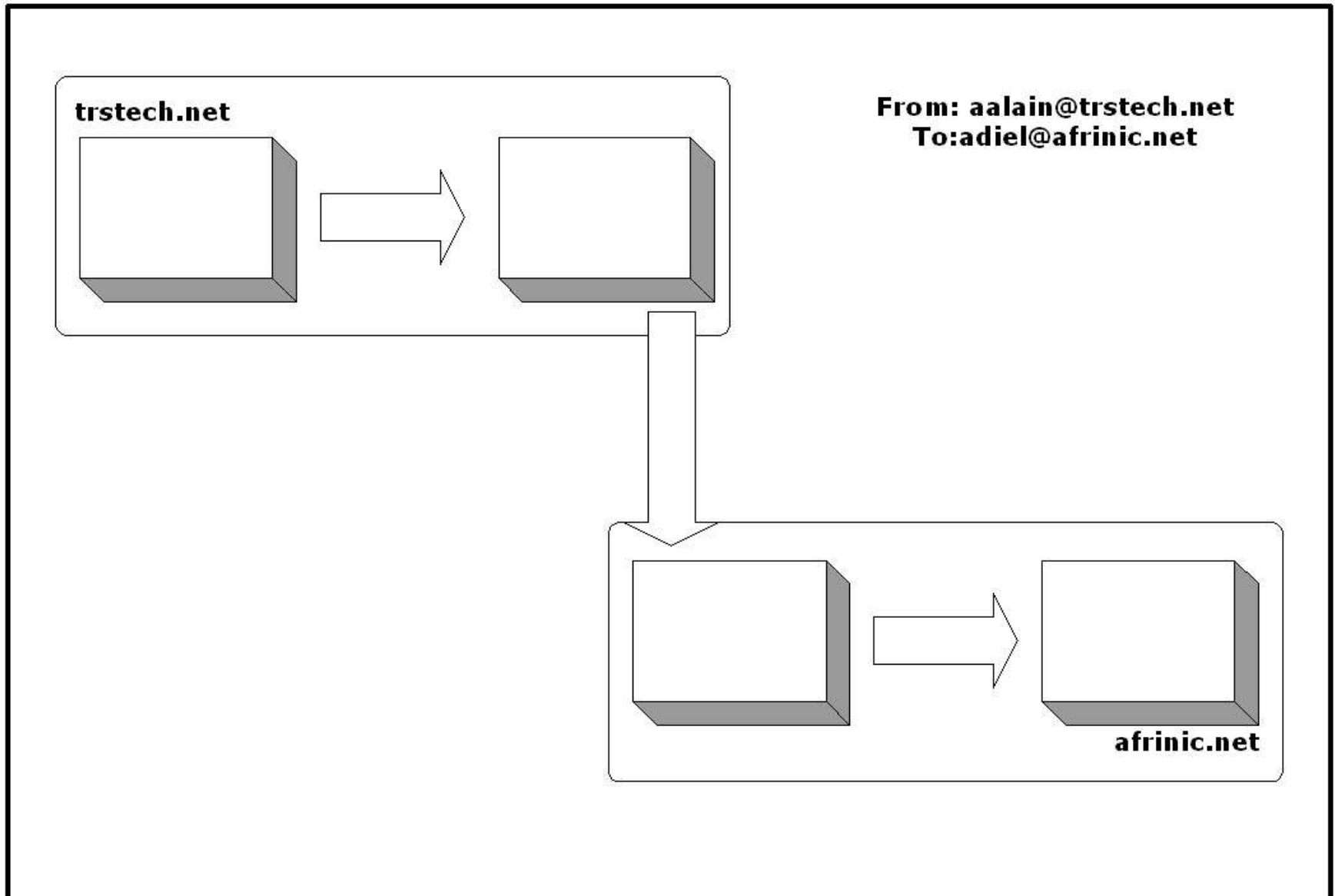
Date: Tue, 4 Apr 2006 18:52:31 -0400

MIME-Version: 1.0

Comment est utilisier SMTP

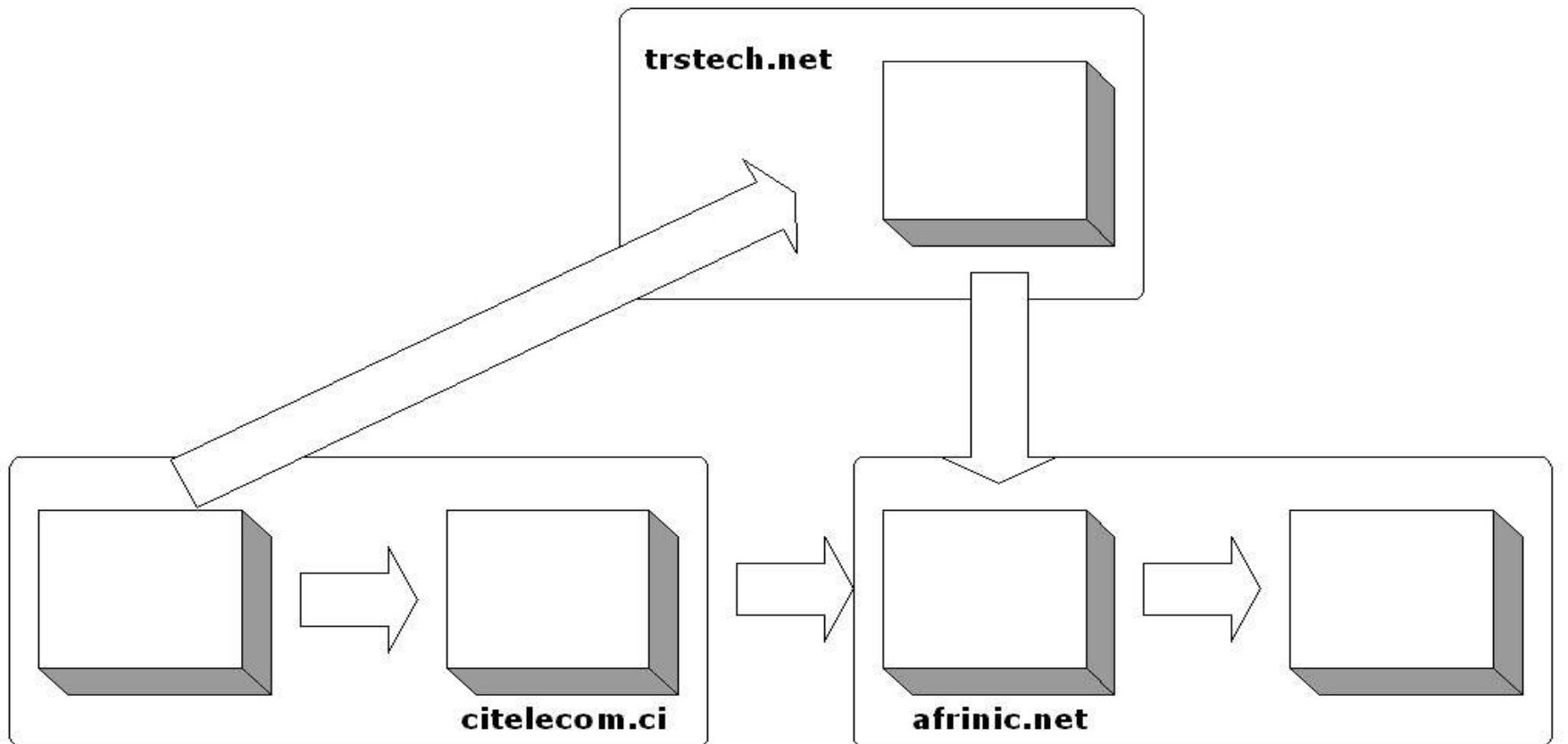


Trafic de base



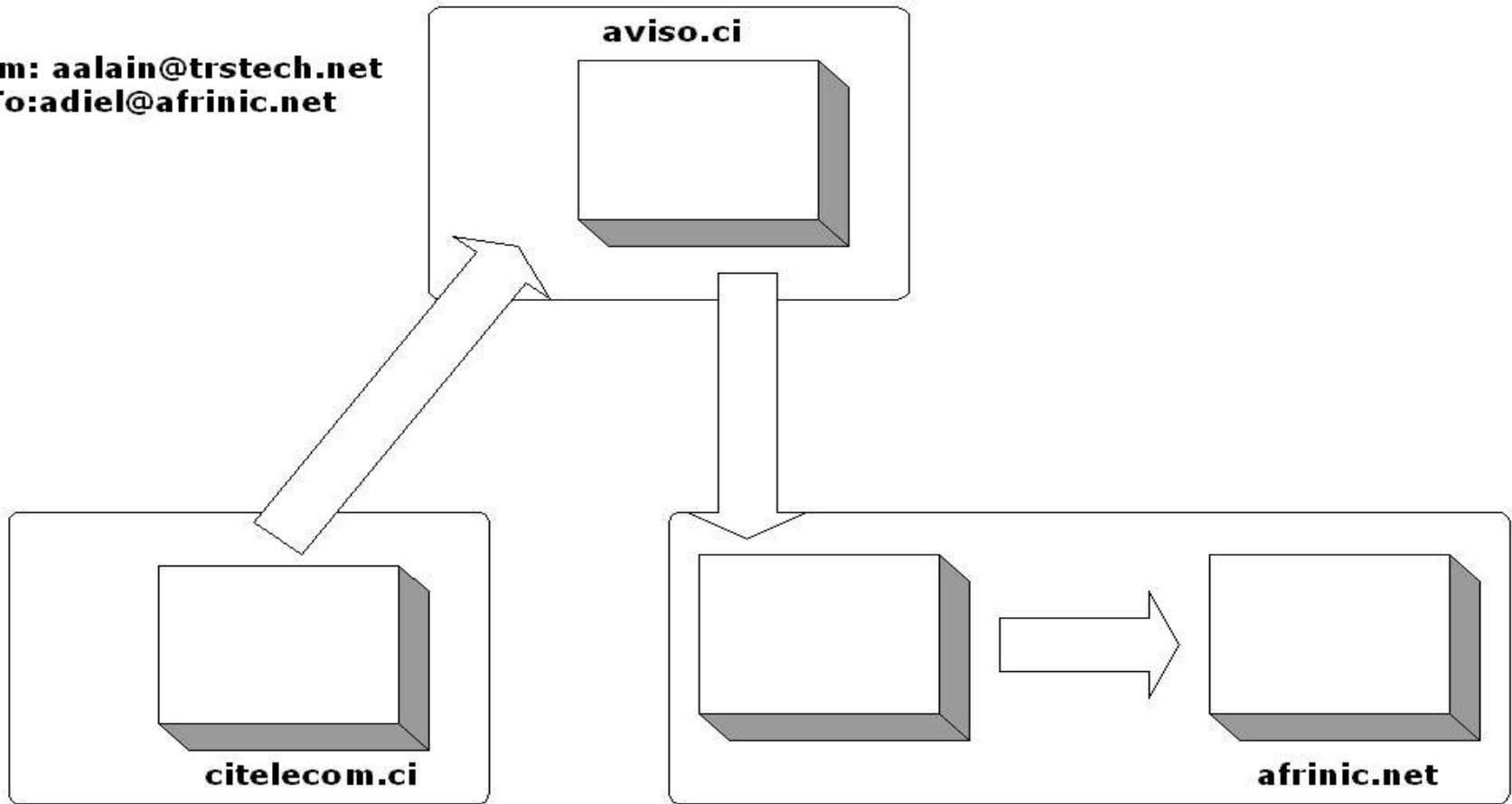
D'un autre domaine

From: aalain@trstech.net
To: adiel@afriNIC.net



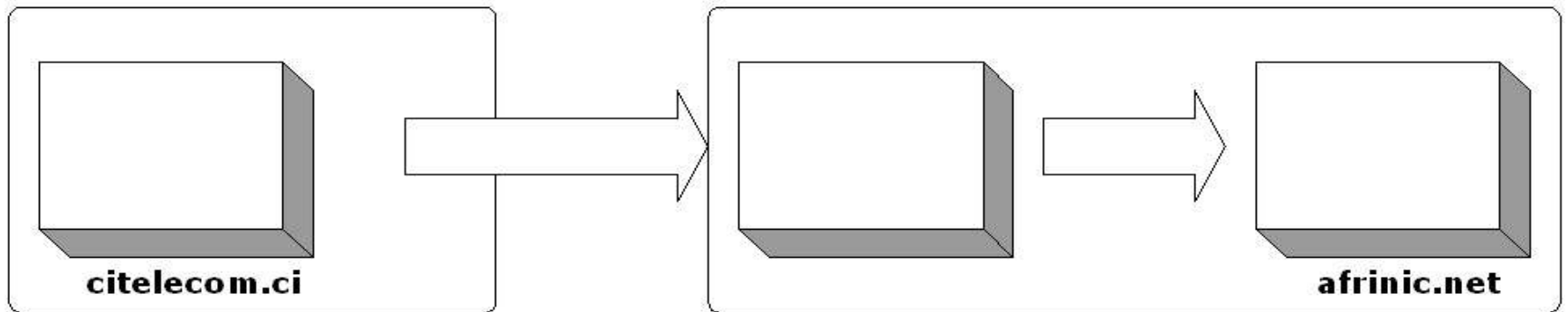
Relais ouvert

From: aalain@trstech.net
To: adiel@afriNIC.net



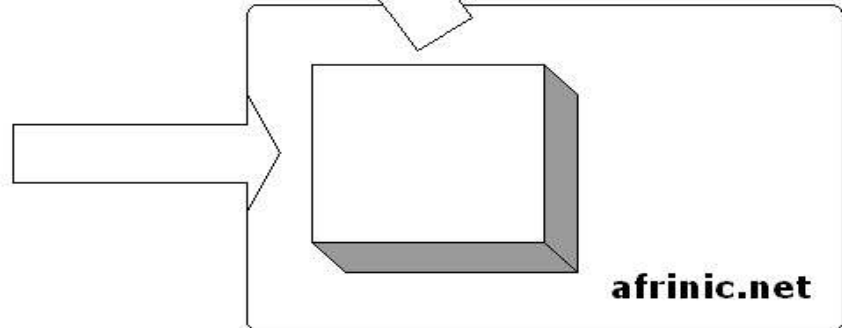
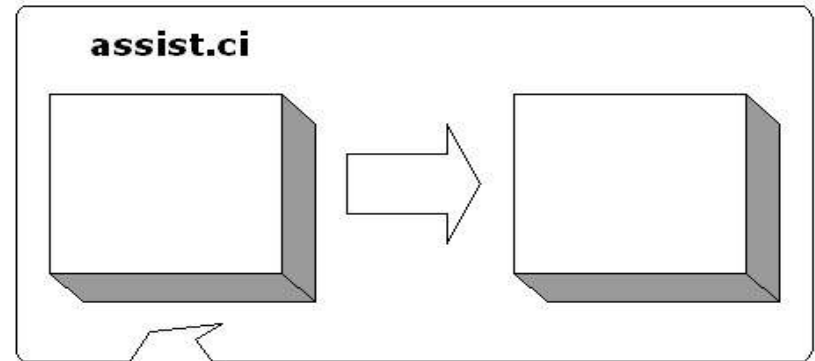
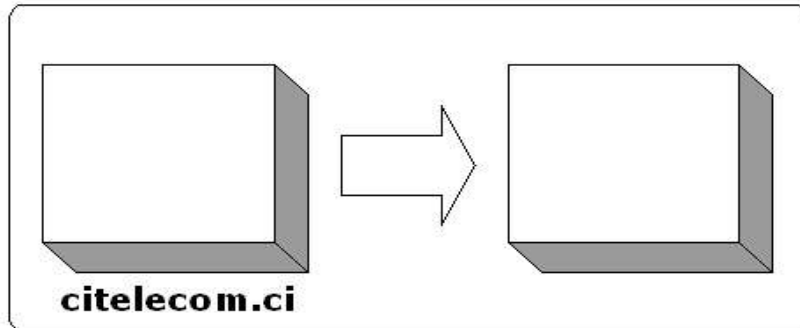
Direct

From: aalain@trstech.net
To: adiel@afriNIC.net



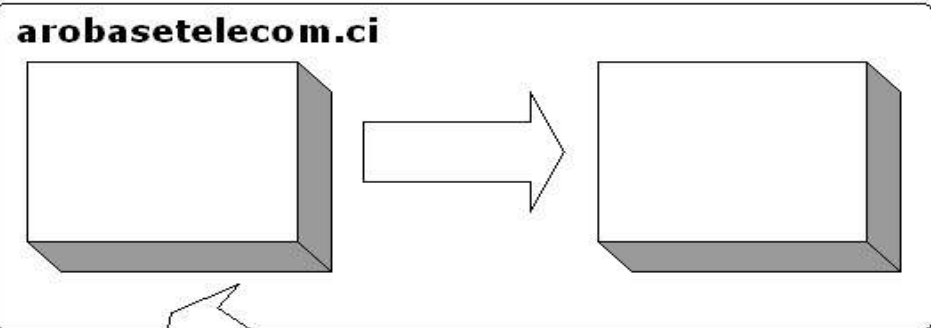
Rebond

From: foo123123@hotmail.com
To: non-existing@afrinic.net
Envelope-From: existing@assist.ci



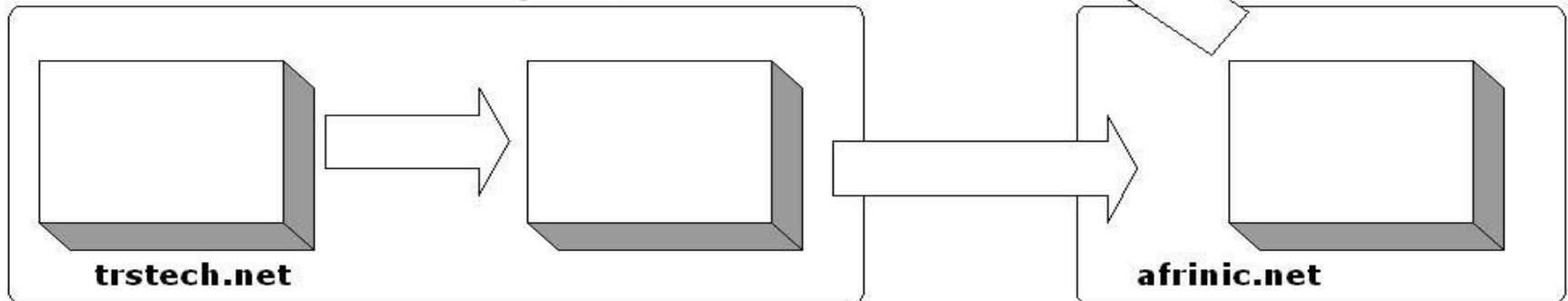
Transfert par MTA

From: aalain@trstech.net
To: adiel@afriNIC.net



Envelope-From: aalain@trstech.net
Envelope-To: adiel@arobasetelecom.ci

Envelope-From: aalain@trstech.net
Envelope-To: adiel@afriNIC.net

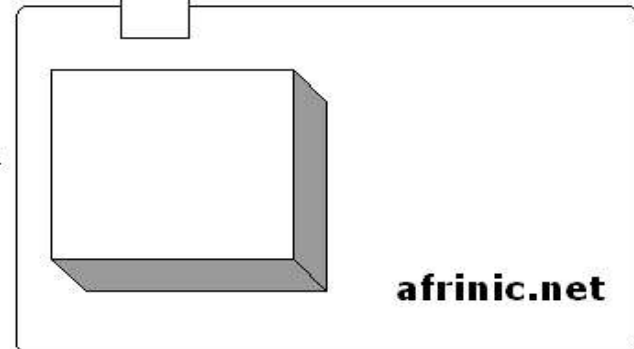
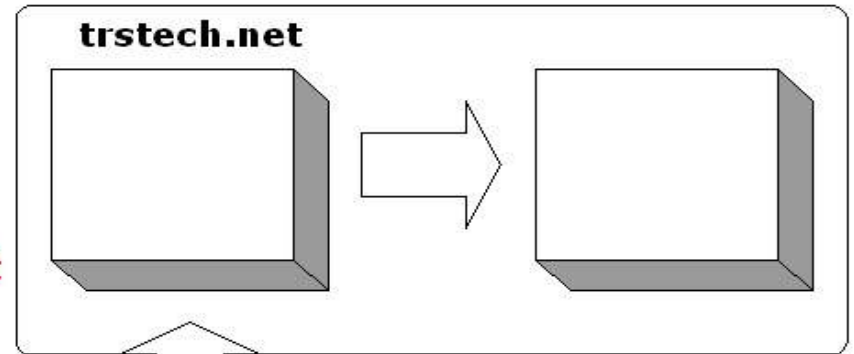
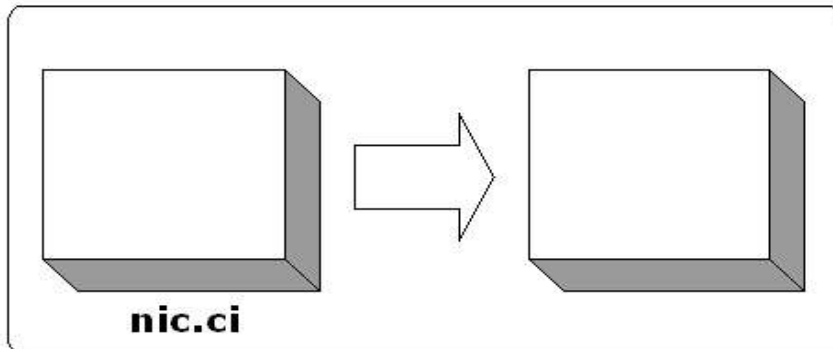


Liste de diffusion

From: paf@nic.ci
To: list@afrinic.net

Envelope-From: list-manager@ripe.net
Envelope-To: paf@trstech.net

Envelope-From: paf@trstech.net
Envelope-To: list@ripe.net



Filtrage

Où peut-on filtrer ?

Sur la machine de l'utilisateur final

- ✓ Chaque utilisateur a un contrôle total
- ✓ Spécialement bon pour le filtrage Bayesian
- ✓ Distribution des coûts de traitement
- x Clients doivent télécharger tous les courriers

Sur le serveur mail du FAI

- ✓ Facile pour les utilisateurs
- ✓ Dans certains cas, les courriers peuvent être rejetés avant la transmission du corps
- ✓ Economie d'espace disque
- x Non flexible pour la configuration des utilisateurs

Problèmes légaux avec le filtrage

- Certains clients peuvent être contre le fait que vous faites des jugements de valeur sur leurs courriers ou, regardez leur contenu
- Alors, Soyez sûr que votre contrat avec le client vous le permet
- Ou permettez aux clients de “opt-in” ou “opt-out” du filtrage
- Le filtrage n'est jamais correct a 100%. Alors soyez sûr que vous n'êtes pas défaillant dans les cas où les filtres prennent de mauvaises décisions

Identification des spams par adresse IP

- Dès la connexion de l'expéditeur, on connaît son adresse IP qui ne peut pas être forgée
- On peut vérifier leur adresse IP contre les listes noires en temps réel
 - Liste noire de plage de IP de spammeurs connus
 - Liste noire de IP de relais/proxy ouverts
- Les “Realtime Blocking Lists (RBLs)” sont consultés via le DNS

Utilisation des RBLs

- Avantages
 - ✓ Facile à configurer
 - ✓ Les requêtes DNS sont relativement rapides et moins coûteuses
 - ✓ D'autres personnes maintiennent la liste
 - ✓ Courriers rejetés avant que le corps ne soit envoyé
- Inconvénients
 - x Les listes vont et viennent (menace de légalité)
 - x N'attrapent pas tous les spams
 - x Inefficace contre les virus et les joe-jobs

Quelles RBLs utilisées ?

- Certaines ne sont pas gratuites
i.e. mail-abuse.org
- Certaines ne sont pas bonnes
Politiques trop draconiennes; Les politiques d'autrui
ne sont pas forcément bonnes pour vous (Le blocage
des IP nigériens n'est pas utile pour un FAI africain)
- Essayer celles qui suivent:
sbl.spamhaus.org (spammeurs connus)
relays.ordb.org (relais ouverts)
bl.spamcop.net (sources dynamiques de spam)

Greylisting(1)

- Entre black-et white-listing, avec une maintenance automatique
- Basé sur le fait que les sources de spams ne se comportent pas comme les systèmes de mail normaux
- Regarde uniquement trois informations sur chaque tentative de délivrance de message:
 1. L'adresse IP de l'expéditeur
 2. L'adresse de l'enveloppe expéditeur
 3. L'adresse de l'enveloppe destinataire

Greylisting(2)

- Si le triplet n'a jamais été vu avant, alors rejeter cette tentative et toutes autres qui viendraient dans une certaine période de temps avec une erreur temporaire

Greylisting(3)

- Avantages

- ✓ Très efficace pour le moment contre les spams/virus
- ✓ Économie de bande passante, de ressources système
- ✓ Maintenance automatique

- Inconvénients

- x Problème d'interopérabilité avec le callback SMTP
- x Problème avec les systèmes à pool de MTA
- x Problème avec les VERP (les listes de diffusion...)
- x Délai parfois important dans l'acheminement des mails
- x Problème avec la gestion du retry de certains MTA
- x Nécessite une base de données

“Whitelists”

- Accepter uniquement des mails des gens que nous connaissons
 - ✓ Efficace pour le blocage des spams
 - ✓ Problème de démarrage (voir prochain slide)
- Actuellement, les spammeurs peuvent forger des messages comme venant des gens que nous connaissons
- Mais pour le moment, ils ne semblent pas collecter des informations sur les gens auxquelles nous sommes associés
- Mais les virus et les « cheval de troie » utilisent souvent les carnets d'adresse locaux

Réception de courriers de gens pas sur la “whitelist”

- Par mot de passe: i.e. s'ils incluent un mot magique dans l'entête objet.
- Par filtrage de contenu: i.e. Avec un score spam faible
- Système Challenge-Réponse
 - ✓ Mettre le courrier dans une queue et envoyer un message
 - ✓ Si l'expéditeur répond, elle est ajoutée à la “liste blanche”.

Réception de courriers de gens pas sur la “whitelist”

- Les Systèmes Challenge-Réponse ne sont pas recommandés
 - x Augmente le problème des spams collatéraux
 - x Mauvaise interaction avec les listes de diffusion
 - x Certains correspondant sont choqués
 - x Difficile à déployer à grande échelle

Inconvénients des "whitelists"

- Difficile/ennuyeux de vous contacter pour la première fois
- Pour une solution serveur
 - Chaque utilisateur a besoin d'une liste blanche séparée et un moyen pour l'éditer
 - L'ajout automatique sur liste blanche des gens auxquelles nous envoyons un courrier n'est pas facile
- Le filtrage à l'étape de "Mail From:" devient plus difficile
 - Enveloppe expéditeur peut être différente du "From:" dans les entêtes

Identification des spams par contenu

- Un humain peut facilement identifier un spam
 - Plus difficile à faire automatiquement
- Rechercher des phrases qui apparaissent souvent dans les spams
- Rechercher des phrases classiques qui n'apparaissent dans les spams
 - Aide à réduire les “false positives”
- Le ratio des deux indique la probabilité des spams
....et avec quelle assurance ?

Inconvénients du filtrage de contenu

- Les spammeurs utilisent diverses astuces pour déguiser leurs actions
 - Codage MIME base64, HTML, division des mots avec des tags invisibles au milieu ... etc
- Un combat permanent
 - Plus les filtres évoluent, les spammeurs changent de tactiques
- Très coûteux en ressources système
- Exposé aux “false positives”
 - A moins que les règles soient définies par les utilisateurs; difficile d'en faire solution niveau serveur

Filtrage “Bayesian”

- Prendre des échantillons de messages connus comme spam ou “non spam” pour construire une table de mots qui apparaissent plus souvent dans l'un que dans l'autre
- Le profile “non spam” est différent pour chacun et par conséquent difficile à deviner par les spammeurs
- Le filtrage est très efficace, mais nécessite une “formation” permanente pour les mails qui passent les filtres’

<http://www.paulgraham.com/spam.html>

Autres moyens d'identification de spams

- Vérification du domaine du MAIL FROM:<...> ou
- Vérification de l'adresse complète par un callback
- Vérification du PTR de l'adresse IP de l'expéditeur
- Vérification de la conformité des courriers aux RFCs...
- Ces règles peuvent attraper certains spams aujourd'hui (jusqu'à ce que les spammeurs s'adaptent). Mais il y a beaucoup de systèmes mal configurés appartenant à des non spammeurs. Vous perdrez des courriers que vous voulez recevoir.

Identification des virus(1)

- Leur volume a beaucoup augmenté ces derniers temps
 - Utilisateurs contents d'ouvrir des fichiers attachés venant d'étrangers!
- Comme les spams, les virus ont une enveloppe expéditeur et des entêtes forgées
- Certains systèmes peuvent bloquer tous les fichiers attachés exécutables
 - Blocage de certaines utilisations légitimes du courrier
 - Certains virus viennent en fichier .zip maintenant

Identification de virus (2)

- Le seul moyen sûr est le filtrage de contenu:
Vérifier les attachements contre les signatures de virus connus
- Certaines solutions sont commerciales, chères, coût augmentant avec le nombre d'utilisateurs
- D'autres sont libres, i.e. clamav
<http://clamav.sourceforge.net/>
- De nouveaux virus sortent tout le temps,
 - Les signatures doivent être mises à jour très régulièrement

"Joe-jobs"

- Spam ou virus envoyés avec une enveloppe expéditeur forgée

MAIL FROM:<innocent@exemple.com>
RCPT TO:<destinataire@domaine-destinataire.com>

- Le message est accepté par les MTA intermediaires et le rebond va à la partie innocente..

Difficultés avec le blocage des "joe-job"

- Tous les rebonds ont une enveloppe expéditeur vide, MAIL FROM:< >
 - Pas utile pour le filtrage
- Les rebonds "Joe-job" sont des rebonds MTA- pas des rebonds des messages que nous avons envoyés
- Le filtrage de contenu pour identifier les rebonds n'est pas utile
- Rejeter tous les rebonds n'est pas une bonne option
 - Utilisateurs se trompent d'adresses
 - Les boîtes sont inaccessibles ou à la limite des quota

Association des rebonds aux messages envoyés

- Malheureusement les rebonds ne sont pas normalisés de manière à permettre cela
- La seule chose que nous savons, c'est que les rebonds vont à l'adresse du “MAIL FROM”
- Alors, une solution est de re-écrire l'adresse “MAIL FROM” à une valeur secrète qui change chaque jour ou à peu près: Connue comme “Variable Envelope Return Path” (VERP)

MAIL FROM:<username=ac7933dc@example.com>

Avantages de VERP

- Les bons rebonds sont gardés, les mauvais sont rejetés
- Un "cookie" cryptographique sera très difficile à imaginer par les spammers
- Dur pour les spammers de collecter les enveloppes expéditeurs
 - Même s'ils les collectent, elles ne sont valables que pendant quelques jours seulement
- Ceci n'est pas une solution contre le spam: C'est une solution "joe-job"... qui tue les spams envoyés avec Mail From: < >

Inconvénients des VERP

- Mauvaise interaction avec les listes de diffusion et les “whitelists” (Si elles regardent le MAIL FROM: au lieu de l'entête From:)
- Les problèmes d'interopérabilité seront minimisés avec la définition d'un standard
 - Plusieurs propositions sont en discussion
- Les utilisateurs doivent envoyer leurs mails, votre MTA
 - Si non les cookies ne seront pas ajoutés et ils perdront des rebonds
- Génère des “nom d'utilisateurs” longs
 - RFC 2821 exige seulement 64 caractères

La communauté et les solutions(1)

- BATV (Bounce Address Tag Validation)
 - Ajout de tag aux « noms d'utilisateurs »
- CSA(Client SMTP authorization)
 - Listes DNS des machines autorisées à envoyer des mails
- SPF(Sender policy Frameworks)
 - Sender-ID de microsoft
 - Liste DNS des machines qui peuvent utiliser une enveloppe expéditeur
 - Bloque complètement le mail forwarding
 - Les allégations sur sa suppression de tous les spams sont exagérées

La communauté et les solutions(2)

- SRS(Sender Rewriting Scheme)
 - Une tentative pour patcher le SPF, mais non sans problème
- Domainkeys(Yahoo!) ou Identified Internet Mail(CISCO)
 - Signer numériquement les messages avec une clé privée par domaine
 - La signature est placée dans l'entête
- Aucune de ses solutions n'est sans problème
 - Certaines ont des problèmes de patente/licence

Que faire(1)?

- Implémenter les RBLs
 - Efficacité surprenante
 - Très facile à faire
 - Faible maintenance
- Implémenter le greylisting
- Considérer l'implémentation des filtres de contenu
 - Pour les utilisateurs qui ont “opt-in”
 - Marquer les spams et laisser les utilisateurs décider
- Penser aux coûts de mise en oeuvre
 - Ces services sont chers à déployer à grande échelle et à gérer
 - Faire payer un coût aux “opt-in” ???

Que faire(2)?

- Sensibiliser les utilisateurs sur les filtres antispam côté client
 - Les filtres bayesian et les whitelists sont plus faciles à ce niveau
 - Trouver ceux qui marchent bien avec client de messagerie
- Eduquer les utilisateurs sur le choix et l'utilisation de leurs adresses électroniques
- Contribuer aux renforcements des lois et de la collaboration nationale et internationale
- Faire son choix (opt-in ou opt-out ?)
- etc.....

La suite ?

- L'IETF va certainement continuer à scruter les solutions de vérification à base du DNS
- L'IRTF va continuer ses travaux sur les spams
 - Que doit être le langage de description de la politique ?
- Devons-nous faire une nouvelle génération de SMTP ?
- Que ferons les spammeurs ?

Questions ??