DE LA RECHERCHE À L'INDUSTRIE

www.cea.fr

Sécurité de la messagerie

Pascal MALTERRE

Direction des applications militaires
Département des sciences de la simulation et de l'information
Cellule technique de sécurité informatique

5 janvier 2015

Summary

Historique

La préhistoire…

- Envoi de messages entre différents utilisateurs des mainframes à la fin des années 60 (MAILBOX)
- Premier envoi d'un message entre deux machines du réseau ARPANET courant 1972 (1 ère utilisation du caractère @)
- En 1973, les mails représentent 3⁄4 du trafic ARPANET – Nombreux problèmes d'interropérabilité (RFC 680, 724, 733, etc.)
- Eric Allman développe delivermail (utilisation de FTP over NCP pour transmettre les messages)

Historique

Les temps anciens (…ou l'avènement de TCP/IP)

- Avant la naissance du protocole SMTP, la transmission des mails se fait au moyen de FTP, UUCP, etc.
- Eric Allman développe sendmail
- La RFC 821 donne les premières spécifications du protocole SMTP (Simple Mail Transfer Protocol) (en 1982)
- La RFC 822 spécifie le format des messages

Historique

Les temps modernes…

- L'évolution vers la messagerie actuelle :
  - Encapsulation de contenus multimédias (MIME)
  - Interconnexion avec les annuaires LDAP, Active Directory
- Les besoins émergents en terme de sécurité :
  - Authentification, confidentialité, etc.
  - Filtrage applicatif : spam, antivirus, etc.

Le mail en quelques chiffres

- 150 milliards de mails par jour ;
- taille moyenne d'un message : 75 Ko (à l'exception du spam) ;
- concernant le spam : 70% du trafic (5 Ko par message en moyenne) ;
- 2,3 milliards de comptes mails ;
- environ 8000 messages stockés par BAL.

Vocabulaire

- MTA (Mail Transfer Agent) : programme responsable de l'acheminement des messages
- MUA (Mail User Agent) : programme permettant à l'utilisateur de lire et d'envoyer des messages
- Dialogue avec le MTA via le protocole SMTP (client uniquement) et avec le serveur de boîtes aux lettres via POP ou IMAP
- MDA (Mail Delivery Agent) : interface entre le MTA et la boîte aux lettres de l'utilisateur

MUA et MDA

- Les MUA sont les programmes utilisés par les utilisateurs pour gérer leur messagerie au quotidien : Microsoft Outlook, Thunderbird, etc.
- Un MDA est généralement directement exécuté par le MTA afin d'ajouter le message à la BAL de l'utilisateur destinataire. C'est en fait le MDA qui prend alors en charge la partie finale du traitement : procmail, maildrop, etc.

Le format des adresses

Une adresse mail est une chaîne de caractère contenant le symbole "@"

- De la forme : "local-part@domain" (la partie "domain" est en fait tout ce qu'il y a après le dernier "@", elle est indépendante de la casse des caractères)
- Caractères ASCII uniquement
- Quelques exemples d'adresses particulières :
  - postmaster@domain.com (RFC 822)
  - abuse@domain.com
  - …

Le format des messages

Un message est constitué de trois parties :

- L'enveloppe est utilisée par le MTA pour l'acheminement du courrier, il s'agit essentiellement de l'expéditeur et du destinataire
- Les en-têtes (ou headers) sont utilisés par les MUA
- Le corps est le contenu du message envoyé au destinataire

Le format des messages

- Le format et l'interprétation des en-têtes sont spécifiés dans la RFC 822 (1982)

« 822 allows much more flexibility than a typical mail- reading program can actually handle ; meanwhile it imposes restrictions that millions of users violate every day »

- Les en-têtes sont de la forme **`Nom: valeur`**

```
Date: Sat, 21 Aug 2004 12:27:41 +0200
X-Spam-Flag: YES
From: Alice <alice@moon.com>
To: Bob <bob@sun.net>
Subject: Fw: failure notice

Wassup?!
...
```

Le routage des mails

Problème

À quel(s) serveur(s) le MTA doit-il s'adresser pour transmettre un message à l'adresse **`bob@example.net`** ?

Le MTA effectue une requête DNS pour trouver les enregistrements MX pour le domaine de l'adresse mail (*i.e.* de tout ce qui est à droite de @)

- en cas d'échec de la requête DNS, cela ne veut pas forcément dire qu'il n'y a pas d'enregistrement MX : le MTA doit réessayer un peu plus tard ;
- s'il y a plusieurs enregistrements MX, alors on utilise le champ priorité pour établir un ordre de préférence ;
- s'il n'existe aucun enregistrement MX pour ce domaine, le MTA doit prendre comme MX l'hôte lui-même.

Summary

Le protocole SMTP

SMTP = *Simple Mail Transfer Protocol*

- Définit la manière de communiquer entre deux MTA (ou entre un MUA et le premier serveur MTA) en utilisant une connexion TCP
- Protocole de type "PUSH"
- Utilise un alphabet ASCII 7 bits (transmission de 8 bits avec le bit de poid fort à 0)
- Le nombre de commandes utilisées est relativement faible (inférieur à 12)
- Le protocole est défini dans la RFC 821 (1982)

La commande HELO

La commande **HELO** permet à la machine source de s'identifier auprès du serveur SMTP

Exemple

`HELO mail.example.org` ie. "Salut, je suis mail.example.org"

- d'après la RFC 1123, le paramètre transmis doit être un nom d'hôte valide au sens DNS ;
- bien que cela soit interdit par la RFC 1123, il existe des implémentations de serveurs SMTP qui refusent des mails suite à des vérifications supplémentaires sur la commande HELO : présence d'enregistrements DNS, cohérence du reverse, cohérence avec le MX, etc.

La commmande MAIL

- La commande `MAIL` permet de spécifier l'adresse de l'expéditeur (« `return-path` ») et de réinitialiser la liste des destinataires
- La commande `MAIL` est suivie de la chaîne "FROM :" et d'une adresse de retour (la chaîne <> représente une adresse vide) et éventuellement des informations supplémentaires (8BITMIME)

```
MAIL FROM: <bob@example.org>
250 ok
```

La commande RCPT

La commande **RCPT** permet d'ajouter une adresse à la liste courante des destinataires

- Cette commande est suivie de la chaîne `TO:` , d'une adresse et éventuellement d'informations optionnelles
- Plusieurs commandes RCPT peuvent être utilisées à la suite les unes des autres

```
RCPT TO :<webmaster@rstack.org> 250 ok RCPT TO :<moutane@rstack.org> 250 ok
```

La commande DATA

La commande DATA permet de transmettre le message

- cette commande n'accepte pas de paramètre ;
- la transmission du message commence immédiatement après la réponse du serveur ;
- chaque ligne du message se termine par <CRLF>, la fin est signalée par la séquence <CRLF>.<CRLF>

```
DATA 354 go ahead Hello world !  .  250 ok 1102890861 qp 5433
```

Commandes obsolètes

- Historiquement, la commande VRFY permettait de vérifier la validité d'une adresse
  - problème de confidentialité vis à vis du Spam ;
  - La RFC 1123 définit un nouveau code d'erreur (252) permettant de traiter la commande VRFY.
- La commande EXPN permet de développer une adresse générique représentant une liste de diffusion
  - le terme "liste de diffusion" peut également désigner un utilisateur local (par exemple : "expn root") ;
  - généralement non implémentée (cf. Spam).

Extended SMTP (ESMTP)

- ces extensions (définies dans la RFC 1425) maintiennent néanmoins une compatibilité ascendante ;
- utilisation de EHLO à la place de HELO, puis code réponse 250

```
bash nc smtp.wanadoo.fr 25
220 mwinf0209.wanadoo.fr ESMTP ***************************
EHLO marley.com
250-mwinf0209.wanadoo.fr
250-PIPELINING
250-SIZE 10485760
250 8BITMIME
```

Remarques

- Lorsqu'un serveur de mail accepte un message, il ajoute un en-tête "Received :" au début du message
- Le protocole TCP n'est utilisé qu'en half-duplex (le client envoie une requête, attend la réponse, etc.)
- Importance de la rapidité de réponse pour les serveurs
- Généralement, les serveurs SMTP ne parsent pas les headers d'un message, excepté pour compter le nombre de saut (nb. de "Received :")

Fiabilité du protocole

Dans le paragraphe 5.3.3 de la RFC 1123, intitulé "Reliable Mail Receipt", il est écrit :

When the receiver-SMTP accepts a piece of mail (by sending a "250 OK" message in response to DATA), it is accepting responsibility for delivering or relaying the message. It must take this responsibility seriously, i.e., it MUST NOT lose the message for frivolous reasons, e.g., because the host later crashes or because of a predictable resource shortage

Objectifs d'un MTA

- accepter les mails issus des processus locaux ;
- accepter les mails provenant d'autres serveurs ;
- transmettre les mails aux processus locaux ;
- transmettre les mails aux autres serveurs.

Architecture d'un MTA

Quelques exemples de MTA

- sendmail (”le” mailer Unix historique) ;
  - approche monolitique, multiples failles de sécurité, configuration complexe, etc.
- qmail
  - architecture modulaire conçue de façon très sécurisée ;
  - configuration aisée, très performant
- Postfix
  - architecture similaire à qmail ;
  - plus moderne (plus d’évolutions possibles).
- OpenSMTPD

Summary

Les formats de BAL

Deux approches :

1. format « opaque » : le contenu de la BAL est complètement masqué pour l'utilisateur ; accès uniquement par un protocole et/ou un client spécifique.
2. format « ouvert » : le contenu de la BAL est accessible par différents MTA/MDA/MUA.

Le format mbox

- un seul fichier est utilisé pour stocker tous les messages présents dans la BAL ;
- Chaque message débute par la chaine « `From` »suivi du message (en-têtes et corps), puis d'une ligne vide ;
- au niveau du système de fichiers, utilisation de verrous pour les accès en écriture à la boîte aux lettres
  - inconvénients : intégrité des messages en cas de crash, plus difficile de mettre en oeuvre un stockage réseau (NFS, etc.)
- format traditionnel sous Unix ;
- plusieurs variantes : `mboxrd`, `mboxo`, `mboxcl`, etc.

Le format `Maildir`

- format popularisé par `qmail` ;
- chaque mail est enregistré dans un fichier, sans aucune modification (conforme à la RFC 822) ;
- Une BAL au format Maildir contient 3 sous- répertoires :
  - `new/` contient les nouveaux messages, ils ne sont pas encore lus par l'utilisateur ;
  - `cur/` contient les messages actuels (déjà lus) ;
  - `tmp/` contient les messages en cours d'élaboration.

- avantages : accès simultanés à la BAL, NFS, etc.
- inconvénients : générer des noms de fichiers uniques.

Généralités

POP *Post Office Protocol*

- RFC 1939
- protocole permettant à un programme local de récupérer les mails stockés dans une BAL distante (ie. présente sur un serveur distant)
- utilise le port `TCP/110`
- Encore majoritairement utilisé par les ISP :
  - protocole est très simple, donc peu d'incompatibilités avec les programmes clients ;
  - le fonctionnement par défaut favorise le contrôle de la taille des boites aux lettres ;
  - convient parfaitement à un accès à Internet non-permanent.

L'authentification POP

Comme beaucoup de protocoles anciens, le mot de passe est transmis en clair par défaut

La commande (optionnelle) APOP

- permet une authentification par challenge/response ;
- le serveur envoie une chaîne contenant (entre-autre) un marqueur de temps dans sa bannière de connexion ;
- De la forme : <marqueur-unique@nom-serveur>
- le client concatène ce marqueur et son mot de passe et renvoie le résultat au serveur qui peut ainsi vérifier son identité.

POP-over-SSL (TCP/995), SASL, etc.

Généralités sur le protocole IMAP

IMAP *Internet Message Access Protocol*
La version courante est la version 4 révision 1 généralement nommée IMAP4rev1 (RFC 3501).

- TCP/143 (ou TCP/993 pour imap over SSL) ;
- Les messages restent sur le serveur, l'utilisation se fait généralement en mode connecté ;
- boîtes aux lettres partagées, accès concurrentiels ;
- création de dossiers IMAP résidant sur le serveur.

Inconvénients du protocole IMAP

La mise en oeuvre du protocole IMAP nécessite des ressources importantes :

- maîtrise de la taille des BAL ?
- au niveau système, certaines fonctionnalités peuvent s'avérer coûteuses (tri, *threading*, etc.),
- au niveau réseau, de multiples connexions IMAP peuvent être ouvertes simultanément lors de l'accès à une BAL ;

La compatibilité avec le protocole peut se révéler très différente entre les divers clients de messagerie

Summary

Généralités sur le Spam

Définition

Courrier électronique non sollicité envoyé le plus souvent massivement

- représente 70% du trafic mail sur internet (plusieurs centaines de millions de mails/jour)
- c'est avant tout un problème économique (le coût d'envoi d'un message est très faible et le stockage des messages est à la charge des intermédiaires et du destinataire)
- à l'heure actuelle, il n'existe aucune solution définitive au problème du Spam
  - une multitude de techniques peuvent néanmoins être utilisées ;
  - le nombre de spams a cessé de croître !

Protections techniques contre le spam

- sur vérifications DNS :
  - validité des enregistrements MX (ou A) pour le domaine de l'adresse mail de l'expéditeur
  - cohérence entre les informations réseaux et les informations du message (par exemple : résolution inverse de l'adresse IP de l'émetteur et comparaison avec le champ From : du message)
- dans les 2 cas, les risques de faux-positifs sont très importants ;
- interropérabilité

RBLDNS

Blacklists IP

- RBL (*Realtime Blackhole Lists*) ;
- Pour chaque mail entrant, rechercher l'adresse IP de l'expéditeur dans des listes publiques d'adresses IP de spammers connus ;
- s'appuient généralement sur le DNS (DNSRBL) ;

- déploiement simple, relativement peu coûteux en terme de ressources CPU, et l'impact au niveau réseau reste faible
- inconvénients : risques importants de faux-positifs, mise à jour

Le *Greylisting*

- technique relativement récente et assez efficace ;
- codes d'erreurs SMTP (RFC 821) :
  - 1yz, 2yz, 3yz : Positive Reply
  - 4yz : Transient Negative Completion Reply
  - 5yz : Permanent Negative Completion Reply
- les sources de Spam ne prennent pas en compte les erreurs SMTP temporaires ;
- fontionne aussi pour les virus de messagerie

Autres : smtp-delay, etc.

Mise en place d'un filtrage applicatif

- recherche de mot-clés spécifiques ;
  - efficace mais facilement contournable, par exemple remplacer le terme Viagra par V1agra
- « *Rule-based scoring systems* »
  - systèmes capables d'éliminer 90% du Spam
  - par exemple : SpamAssassin
- filtrage basé sur des méthodes probabilistes (filtrage bayésien)

Principe

- Le responsable d'un domaine de messagerie publie dans le DNS (RR de type `TXT`) la liste des machines de son domaine (les "serveurs de courrier sortant") autorisées à envoyer des mails
- Lors de la réception du message, le `MTA` peut vérifier s'il y a correspondance avec la valeur de la commande `MAIL FROM` (`Return-path`)
  - la vérification peut également être effectuée au niveau du `MUA` si ce dernier dispose à ce moment de la valeur du `Return-path`

(pros) Lutte contre l'usurpation de l'expéditeur

(cons) Redirections, *return-path* vides, nomadisme, etc.

DKIM

Domain Keys Identified Mail

Principe

- Ajout d'un entête contenant une signature cryptographique sur une partie du message (corps + certains entêtes)
- La signature peut être vérifiée (la clé publique est généralement stockée dans un enregistrement DNS)

(pros) Transparent par rapport au routage des mails, lutte contre le *phishing*

(cons) L'enveloppe n'est pas prise en compte dans la signature, pas de protection contre le rejeu, modification en cours de route du message possible

Commissariat à l'énergie atomique et aux énergies alternatives DAM/DSSI
Centre de Bruyères-le-Châtel | 91297 Arpajon Cedex
T. +33 (0)1 69 26 40 00 | F. +33 (0)1 69 26 40 00
Établissement public à caractère industriel et commercial
RCS Paris B 775 685 019