

Sender-ID, SPF et leur famille : authentifier l'émetteur de courrier via le DNS

Stéphane Bortzmeyer

<bortzmeyer@nic.fr>

Représentant l'AFNIC au groupe de contact contre le *spam*

14 septembre 2004

Résumé

On sait que le système de courrier électronique sur Internet et notamment son protocole Simple Mail Transfer Protocol (SMTP) n'authentifie aucune des informations reçues depuis un autre serveur de courrier. Cette absence d'authentification facilite la tâche des *spammeurs* ou bien des usurpateurs d'identité¹.

Une famille de protocoles, comme Sender Policy Framework (SPF) et Sender IDentification (Sender-ID), a été développée pour permettre une authentification des émetteurs de courrier, en publiant dans le Domain Name System (DNS) la liste des serveurs autorisés pour un domaine.

Ces protocoles sont d'ores et déjà en cours de déploiement et sont susceptibles de modifier profondément le fonctionnement du courrier électronique.

Cet exposé explique le fonctionnement de ces protocoles, pour un public non technique.

1 Rappels

Le protocole utilisé pour le transport du courrier sur Internet, SMTP, normalisé dans le RFC 2821², a un défaut : il fait entière confiance à l'émetteur du message quant à l'authenticité des adresses d'expédition. Avant de s'exclamer que les auteurs de SMTP étaient vraiment imprudents, il faut se rappeler que c'est l'architecture qu'ils ont conçu qui a assuré le succès de l'Internet et le développement formidable de la messagerie électronique, alors que des normes peut-être plus sûres, mais extrêmement rigides, n'ont jamais été déployées et donc évidemment jamais utilisées pour le *spam*.

Le fonctionnement du courrier sur Internet repose sur trois piliers :

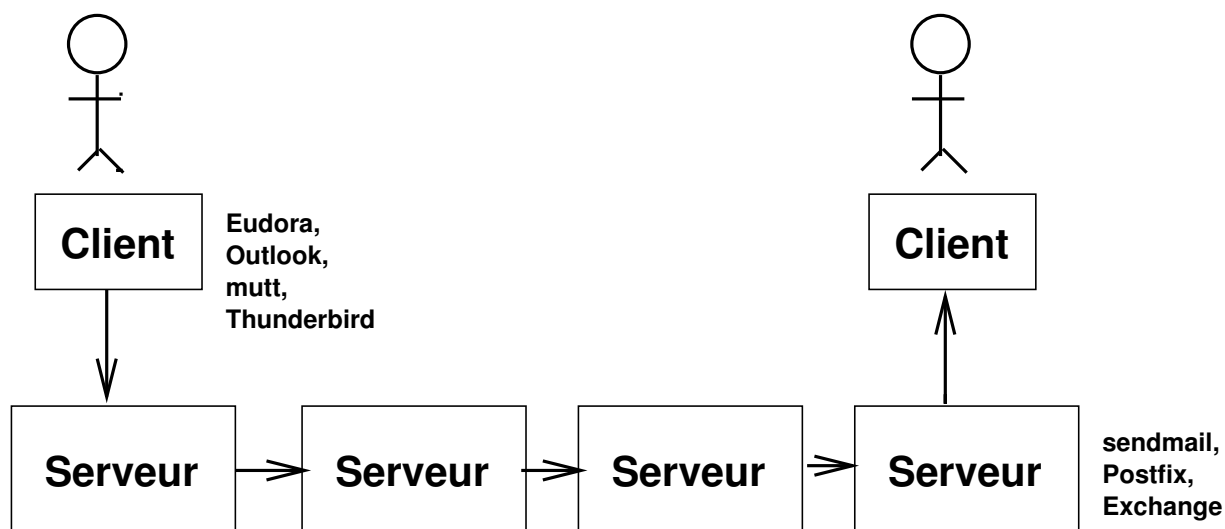
1. Une architecture (qui n'a jamais été décrite par écrit) où il n'y a pas d'arrangement préalable entre les parties, ni d'autorité centralisée. Chaque serveur de courrier envoie potentiellement à chaque autre serveur.
2. Un protocole, SMTP
3. Un format des messages, souvent connu par le numéro du Request For Comments (RFC), 2822, qui spécifie un certain nombre d'en-têtes au message (`From`, qui indique l'expéditeur, `Date`, et des en-têtes plus techniques, destinés au débogage et à l'enquête, comme `Received`, qui indique les serveurs traversés).

1.1 Architecture

Il est courant que plusieurs serveurs successifs, appartenant à des organisations différentes, soient utilisés pour que le message atteigne sa destination. La figure 1 montre un tel cas.

¹Ce qu'on nomme le *phising*

²Pour voir le RFC de numéro NNN, <http://www.ietf.org/rfc/rfcNNN.txt>, par exemple <http://www.ietf.org/rfc/rfc2821.txt>



La flèche indique le sens de circulation du message, le protocole SMTP est utilisé entre les serveurs.

FIG. 1 – Architecture du courrier électronique

```

Récepteur: 220 afnic.fr ESMTP server ready
Émetteur : EHLO mail.u-paris.fr
Récepteur: 250-afnic.fr
Récepteur: 250 SIZE
Émetteur : MAIL FROM:<jean@example.fr>
Récepteur: 250 <jean@example.fr> sender ok
Émetteur : RCPT TO:<jerome@afnic.fr>
Récepteur: 250 <jerome@afnic.fr> recipient ok
Émetteur : DATA
Récepteur: 354 okay, send message
Émetteur : Resent-From: jerome@u-paris.fr
Émetteur : (Autres en-têtes puis corps du message)

```

FIG. 2 – Une session SMTP

1.2 Protocole

Lors d'une session SMTP entre deux serveurs, l'appelant transmet à l'appelé l'adresse de l'expéditeur. C'est ce qu'on appelle l'enveloppe (en anglais, *enveloppe from* ou bien MAIL FROM ou encore 2821 FROM, du numéro du RFC).

La figure 2 montre une session SMTP typique, qui correspond au scénario 1.4.3. On y voit la transmission de l'enveloppe (notamment le MAIL FROM) et le tout début du message.

1.3 Format

Voici (figure 3) un message tel qu'il apparaît sur le réseau. Naturellement, la plupart des logiciels de courrier le présentent de manière plus adaptée à l'utilisateur.

Le message est séparé en deux parties, les en-têtes, structurés selon le RFC 2822 et un corps qui est typiquement du texte (c'est la partie après la ligne vide).

On ne voit pas l'enveloppe ici, elle est gérée par SMTP (voir section 1.2) et typiquement jamais affichée, comme si votre courrier papier était toujours ouvert avant de vous être transmis.

Si on suspecte un faux, il faut analyser ces en-têtes, qui sont souvent mal documentés et faire appel à diverses bases de données en ligne comme celles des Regional Internet Registries (RIR). C'est actuellement un travail d'expert que le destinataire ne peut pas faire seul.

```
Received: from localhost (net-206-66.noicomnet.it [194.153.206.66])
        by rs2.bips.noicomnet.it (8.11.6/8.11.6) with SMTP id i8A97Rr24581;
        Fri, 10 Sep 2004 11:07:27 +0200
Subject: Re: [governance] selection process
From: Vittorio Bertola <vb@bertola.eu.org>
To: avri@acm.org
Cc: governance@lists.cpsr.org
Date: Fri, 10 Sep 2004 11:06:31 +0200
```

If we can make it happen in time, I agree that this could be a good solution. But in this case ...

FIG. 3 – Un message

1.4 Identités

Les sections précédentes ont montré que la question de l'authentification allait être un peu plus complexe que prévu. En effet, il existe plusieurs identités lors de l'envoi et de la réception d'un message et il n'est pas évident de savoir laquelle authentifier. Voyons quelques scénarios.

1.4.1 Le cas le plus courant

jean@example.fr écrit à nicole@laposte.net. C'est le cas le plus classique et le plus simple. Les serveurs de messagerie de la Poste recevront le message depuis les serveurs de messagerie d'Example. Les différentes identités (From de l'en-tête, MAIL FROM de l'enveloppe) seront les mêmes, jean@example.fr.

1.4.2 Les listes de diffusion

Maintenant, jean@example.fr écrit à la liste de diffusion cheval@cru.fr, où se retrouvent les amateurs de chevaux, parmi lesquels nicole@laposte.net. Lorsque le message sera reçu par les serveurs de messagerie de la Poste, le From de l'en-tête vaudra toujours jean@example.fr alors que le MAIL FROM de l'enveloppe vaudra quelque chose comme cheval-list-admin@cru.fr. Les identités seront différentes.

1.4.3 Suivi automatique

Inlassable, jean@example.fr écrit à jerome@u-paris.fr. Mais ce dernier a quitté l'Université et travaille désormais à l'AFNIC. Il a mis en place un suivi automatique du courrier³ vers jerome@afnic.fr et le message arrivera donc sur les serveurs de l'AFNIC avec un MAIL FROM de l'enveloppe inchangé, jean@example.fr, mais sans venir des serveurs d'Example, puisque le message a été retransmis par les serveurs de l'Université.

1.4.4 Un travailleur nomade

jean@example.fr est à un congrès à Washington. Il envoie du courrier depuis l'hôtel en utilisant le serveur de messagerie de l'hôtel. Le MAIL FROM de l'enveloppe sera fixé par l'hôtel, par exemple guest-services@hotel.com et le From de l'en-tête sera jean@example.fr.

1.4.5 Envoyez donc cet article à un ami

jean@example.fr lit un article sur le site Web du Quotidien du Cheval et il veut le transmettre via le classique bouton "Envoyez à un ami". Lorsqu'il le fait, on lui demande son adresse de courrier. Le message partira avec From de l'en-tête à jean@example.fr et un champ Sender de l'en-tête à webmaster@cheval.fr, qui identifiera le site Web⁴.

³Par exemple par le biais d'un fichier .forward sur une machine Unix

⁴Dans la réalité, les sites comme celui du Quotidien du Cheval utilisent des techniques très variées pour gérer ce problème, en raison de l'absence de règles claires. La technique présentée ici avec le champ Sender n'est qu'une des techniques possibles.

```
Received: from (146.82.138.7) [211.216.136.75]
    by master.debian.org with smtp (Exim 3.35 1 (Debian))
    id 1C5ATe-0005Fu-00; Wed, 08 Sep 2004 16:59:43 -0500
Received: from (HELO xpv) [99.148.158.201]
    by 146.82.138.7 with ESMTP id 48155C866E2;
    Wed, 08 Sep 2004 22:58:44 +0100
Message-ID: <n2a--0-49s79c@bk05.rpe4>
From: <anfydlqpxm@amazon.com>
Subject: (Un texte en coréen)
To: benoit@debian.org
Date: Wed, 08 Sep 04 22:58:44 GMT
X-Mailer: Microsoft Outlook Express 5.50.4522.1200
```

FIG. 4 – Un spam prétendument envoyé depuis Amazon

1.5 Jusqu'ici, tout allait bien

Dans tous les scénarios cités plus haut, le serveur émetteur pouvait mettre ce qu'il voulait, aucune authentification n'étant effectuée.

Mais aujourd'hui, cette absence d'authentification est largement utilisée par des méchants. Citons deux exemples courants, un *spammer* et un escroc.

1.5.1 Les délinquants se masquent

La sensibilité (justifiée) des utilisateurs au problème du *spam* est telle que la plupart des *spammers* mentent sur leur identité. Soit ils veulent échapper aux poursuites judiciaires, si leur action est illégale, soit ils veulent passer les filtres qu'ont mis en place les Fournisseurs d'Accès Internet (FAI). Si on sait que "amazon.com" est dans beaucoup de listes blanches, il est tentant de se faire passer pour Amazon, c'est ce que l'on voit sur la figure 4. Le nom de domaine dans le From est mensonger⁵.

Ces mensonges garantissent ainsi au *spammer* un quasi-anonymat.

1.5.2 Ayez confiance, tapez votre mot de passe

Une autre forme d'usurpation d'identité se produit lorsque l'expéditeur tente de vous faire croire que le message vient d'une grande institution respectable, une banque, par exemple, et va tenter d'obtenir de vous des informations confidentielles. C'est le *phishing*. Typiquement, le message semble envoyé depuis "serious-bank.com", la page Web qu'il indique ressemble à celle de Big Bank et le mot de passe de votre compte que vous taperez atterrira chez l'usurpateur. La figure 5 montre un tel exemple, où la première victime est le site de vente en ligne eBay. Une astuce HTML permet de déguiser l'adresse du lien pour ceux qui regardent le message avec un logiciel de courrier qui comprend HTML.

1.5.3 Aux grands maux, les grands remèdes

Le problème du *spam* et celui du *phishing* sont tels que beaucoup d'acteurs sont prêts à casser des œufs pour faire l'omelette, selon un cliché qui revient très fréquemment dans les discussions sur la lutte anti-spam. Ils sont prêts à imposer une authentification des domaines émetteurs. Avant de voir SPF et Sender-ID, voyons les autres solutions.

2 Solutions plus anciennes

L'authentification des individus émetteurs existe depuis très longtemps, via des techniques comme Pretty Good Privacy (PGP), normalisée dans le RFC 2440. Très efficace, elle est utilisée par tous ceux qui veulent faire des annonces incontestables (découverte d'une faille de sécurité

⁵Le message de la figure 4 comprend d'autres mensonges comme le faux numéro de version d'Outlook, que SpamAssassin a su détecter. L'utilisation d'Outlook fait penser à un utilisateur individuel alors que les *spams* sont typiquement envoyés par des logiciels spécialisés.

Received: from s9009.hostcentric.net (s9009.hostcentric.net [66.40.7.4])
by mx1.nic.fr (Postfix) with ESMTP id 288D094005
for <nic@nic.fr>; Mon, 6 Sep 2004 13:46:33 +0200 (CEST)
Received: (qmail 18974 invoked by uid 48); 6 Sep 2004 11:36:31 -0000
Date: 6 Sep 2004 11:36:31 -0000
To: nic@nic.fr
Subject: Security Measures (SafeHarbor) (KMM05092004618V76837L0KM)
From: CustomerSupport@eBay.com

Dear eBay member

We recently noticed one or more attempts to log in to your eBay account from a foreign IP address and we have reasons to believe that your account was hijacked by a third party without your authorization. If you recently accessed your account while traveling, the unusual log in attempts may have been initiated by you. However, if you are the rightful holder of the account, click on the link below, fill the form and then submit as we try to verify your identity. <http://66.132.227.169/e/a/signin.ebay.com/aw-cgi/index2.htm>

FIG. 5 – Une usurpation : le message est prétendument envoyé depuis eBay

par un Computer Emergency Response Team (CERT), par exemple). Mais elle n'a jamais décollé dans le grand public.

Toutes les techniques bâties sur la cryptographie ont connu le même sort : trop lourdes, trop difficiles pour l'utilisateur et nécessitant parfois des infrastructures chères et complexes comme les Public Key Infrastructure (PKI).

3 Authentification dans le DNS

On envisage donc aujourd'hui un mécanisme plus léger : authentifier le domaine et non pas l'individu, et le faire en publiant, grâce au DNS, la liste des serveurs de messagerie autorisés à envoyer du courrier pour ce domaine.

Il existe plusieurs propositions techniques, qui sont apparemment toutes en train de converger. Le principe est simple (figure 6) : un serveur SMTP est à peu près sûr de l'adresse IP de l'autre serveur SMTP qui est en face. Elle est très difficile à falsifier. Il extrait donc le domaine de l'adresse d'expéditeur et regarde dans le DNS les serveurs SMTP de ce domaine. Si le serveur qui tente de lui envoyer un message est dans cette liste, le message est accepté sinon il est, au choix, refusé, détruit ou soumis à des tests plus approfondis.

Ainsi, on authentifie le dernier canal de communication utilisé (entre les deux derniers serveurs), pas le message lui-même comme le faisait PGP. Mais on espère que la facilité de déduction compensera la réduction des fonctions.

Une fois posé ce principe très simple, d'innombrables détails viennent compliquer la tâche. Le principal est la question de l'identité à tester (exposée en section 1.4). Que veut dire "Je représente smith@example.com" ?

3.1 SPF

Le plus connu de ces nouveaux protocoles, et de loin le plus répandu, est SPF, développé par Pobox. SPF teste uniquement le MAIL FROM de l'enveloppe.

SPF brise notamment le suivi automatique du courrier (scénario de la section 1.4.3).

SPF est largement déployé dans le monde. Le site Web de SPF propose documentations et aide.

3.2 Sender-ID

Aujourd'hui, Sender-ID est en discussion au groupe de travail MTA Authorization Records in DNS (MARID), de l'Internet Engineering Task Force (IETF), qui veille à ce que la normalisation couronne un standard neutre, qui ne soit pas sous la menace de brevets léonins et qui ne dépende pas pour son évolution d'une seule entreprise. Après de chaudes discussions, MARID a adopté une stratégie qui ressemble beaucoup à celle de SPF.

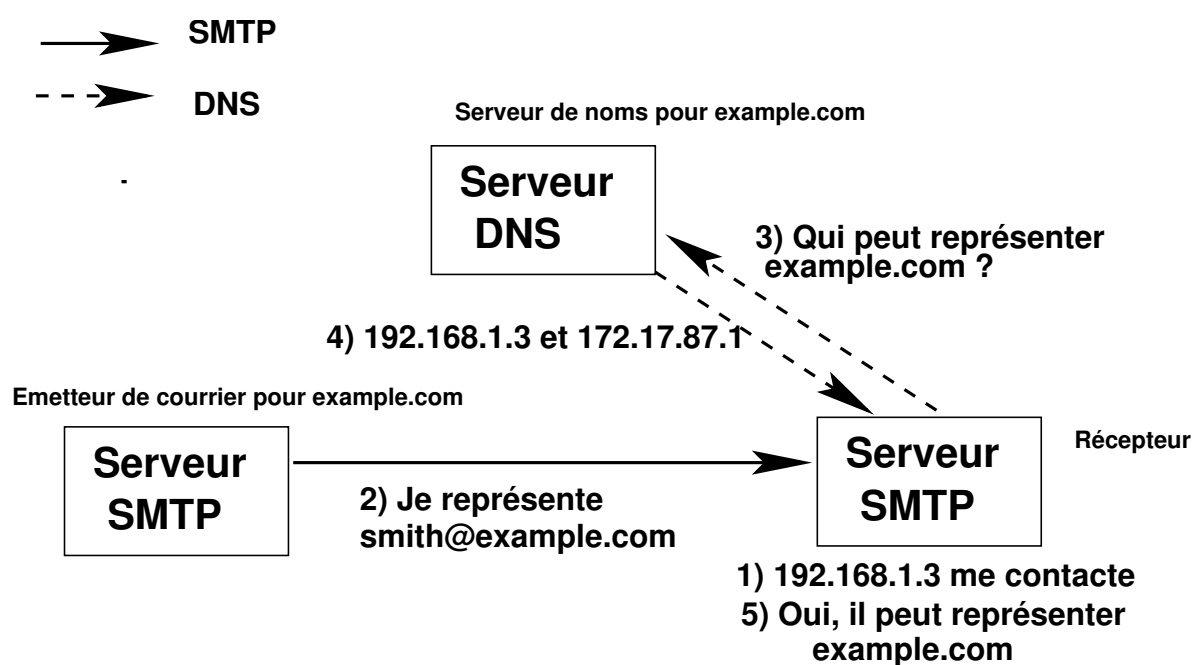


FIG. 6 – Principe de l'authentification dans le DNS ; les numéros indiquent l'ordre des opérations

Sender-ID offrira le choix de l'identité à tester. Pour l'instant, il semble qu'il pourra tester le MAIL FROM de l'enveloppe ou bien le Purported Responsible Address (PRA), extrait des en-têtes⁶ et qui sera en général le From de l'en-tête.

4 Résultats et conséquences pratiques

SPF (ou ses semblables) limite le *spam* ? Indirectement, seulement : un *spammer* peut mettre des enregistrements DNS dans son domaine. Mais SPF contribue à assécher le marécage : si tous les *spammers* sont obligés d'agir à visage découvert, les vocations devraient se tarir.

Dans la tradition du DNS, outil décentralisé, les décisions seront prises par le gérant du domaine ainsi que par celui du serveur de messagerie, qui décide ou pas d'activer les tests SPF ou Sender-ID.

Il ne faut pas se faire d'illusions sur la possibilité de solutions techniques invisibles et sans douleur. Toutes les méthodes anti-spam ont un coût, parfois très élevé. De même qu'il est agaçant de devoir fermer sa porte à clé et d'emporter ses clés, de même l'Internet de demain sera probablement moins pratique et plus contraignant.

En outre, la transition sera douloureuse : lorsque l'usage de SPF deviendra suffisamment important pour que des sites refusent tout message venant de domaines non-SPF, beaucoup de messages seront perdus, victimes innocentes de la guerre mondiale contre le spam. Ainsi, dans un scénario comme celui de la section 1.4.5, des messages légitimes pourront être refusés car provenant d'un serveur non autorisé. De même, les travailleurs nomades comme celui de la section 1.4.4 vont avoir des difficultés. Un gros travail de déploiement et de support est donc à prévoir.

Quel est l'avenir de SPF ? A l'origine, ce n'était qu'une proposition parmi d'autres, portée par une entreprise privée. Seule de ces propositions, SPF a réussi à entraîner l'adhésion de beaucoup, par un gros effort de marketing mais aussi par la mise à la disposition de tous d'excellentes documentations et de logiciels libres mettant en œuvre SPF.

Aujourd'hui, deux propositions se font face, le SPF traditionnel, limité mais déjà mis en œuvre et déployé et Sender-ID qui, à l'heure où ces lignes sont écrites, n'a pas encore été fi-

⁶C'est l'algorithme - absolument trivial - d'extraction de l'adresse qui fait l'objet d'une demande de brevet de Microsoft. Une licence - très restrictive - est délivrée par Microsoft pour exploiter cet algorithme.

nalisé. Le choix entre eux, ainsi que la décision de suivre cette voie de l'authentification ou pas, dépend désormais des acteurs de l'Internet.

5 Glossaire

AFNIC Association Française pour le Nommage Internet en Coopération. Registre des Country-Code Top Level Domain (ccTLD) "fr" et "re". <http://www.afnic.fr/>

ccTLD Country-Code Top Level Domain. Un domaine national comme "fr" en France ou bien "de" en Allemagne.

CERT Computer Emergency Response Team. Un centre de compétences sur la sécurité informatique, dont une des tâches est de publier des alertes de sécurité avec les correctifs suggérés. Ces alertes doivent être authentifiées et les CERT utilisent PGP.

DNS Domain Name System. Le mécanisme d'annuaire distribué de l'Internet, qui permet, non seulement de traduire les noms en adresses IP, son utilisation la plus connue, mais aussi de récupérer diverses informations comme la liste des serveurs de courrier autorisés pour le domaine, ce que font SPF et Sender-ID.

FAI Fournisseur d'Accès Internet. Comme Free, No-Log ou Nerim en France.

IETF Internet Engineering Task Force. Organisme chargé de la normalisation des protocoles Internet. Le groupe MARID, en fait partie. <http://www.ietf.org/>.

MARID MTA Authorization Records in DNS. Le groupe de travail de l'IETF qui tente de normaliser les protocoles d'authentification des serveurs de courrier via le DNS. <http://www.ietf.org/html.charters/marid-charter.html>

NIC Network Information Center. À l'origine un organisme chargé d'assurer l'information sur l'Internet et d'attribuer des ressources uniques comme les adresses IP. Désigne désormais un organisme chargé de gérer un domaine comme l'Association Française pour le Nommage Internet en Coopération (AFNIC) gère "fr" ou bien le DENIC gère "de". On dit aussi un registre.

PGP Pretty Good Privacy. Un système de signature électronique par la cryptographie très sûr et très utilisé, mais dans des communautés restreintes, typiquement dans le milieu de la haute technologie.

PKI Public Key Infrastructure. Infrastructure de gestion de clés publiques pour la cryptographie. Typiquement un logiciel et un ensemble de procédures pour créer des clés, les distribuer, les annuler, etc.

PRA Purported Responsible Address. Un algorithme trivial (mais apparemment breveté) pour extraire des en-têtes du message l'adresse du responsable de la dernière injection du message dans le système de courrier. Dans les cas simples comme celui en 1.4.1, c'est seulement le From de l'en-tête. Dans un cas comme celui en 1.4.5, cela serait le Sender de l'en-tête.

RFC Request For Comments. Les "textes sacrés" de l'Internet. Certains ont un caractère normatif mais pas tous. On les identifie par un numéro par exemple 1034 et 1035 pour le DNS. Ils sont tous accessibles publiquement et redistribuables librement.

RIR Regional Internet Registries. Organismes comme le RIPE-NCC en Europe chargés d'attribuer les adresses IP.

Sender-ID Sender IDentification. Un protocole en cours de normalisation à l'IETF pour authentifier les serveurs de courrier via le DNS.

SMTP Simple Mail Transfer Protocol. Le protocole de très loin le plus courant dans l'Internet pour la communication entre serveurs de messagerie. Normalisé dans le RFC 2821.

SPF Sender Policy Framework. La seule technique d'authentification des serveurs de courrier par le DNS qui aie été largement déployée. Développé par Pobox. <http://spf.pobox.com>