

Les normes du courrier électronique enfin entièrement internationalisées

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 7 septembre 2008. Dernière mise à jour le 12 mars 2013

<https://www.bortzmeyer.org/courrier-entierement-internationalise.html>

Parmi les protocoles utilisés sur Internet, presque tous sont internationalisés depuis longtemps. Le dernier gros manque concernait les adresses de courrier électronique, obligées de s'en tenir à `stephane@internet-en-coopération.fr` alors qu'on voudrait pouvoir écrire à `stéphane@internet-en-coopération.fr`. Désormais, nous avons une solution normalisée, "*Internationalized Email*", dont les spécifications viennent d'accéder au statut de norme IETF.

Bien sûr, pour l'exemple ci-dessus, n'utilisant que l'alphabet latin, le gain n'est pas évident. Mais il l'est beaucoup plus si vous voulez écrire votre adresse avec l'écriture arabe ou chinoise.

Le **contenu** des courriers était internationalisé depuis longtemps, au moins depuis MIME (RFC 1341¹ à l'origine, en juin 1992). Mais les adresses ne l'étaient pas encore. Une solution était en développement depuis longtemps à l'IETF, sous le nom d'EAI ("*Email Addresses Internationalization*") et avait été publiée à l'origine en 2008, avec un statut « expérimental ». Elle est désormais, après une histoire complexe, entièrement normalisée.

Il y a deux parties à l'adresse (décrites dans le RFC 5322) : la partie locale, à gauche du @ et le nom de domaine à droite. Si ce nom de domaine, depuis la sortie du RFC 3490, en mars 2003, peut s'écrire en Unicode (norme IDN), le courrier exigeait toujours un nom de domaine en ASCII donc un message de non-délivrance, par exemple, allait afficher la forme Punycode du nom. Ce n'était pas une vraie internationalisation.

La nouvelle norme, qui réalise enfin cette internationalisation, est spécifiée dans les RFC suivants :
— RFC 6530, "*Overview and Framework for Internationalized Email*" décrit le cadre général. Si on ne lit qu'un seul RFC de la série, c'est celui-là.

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc1341.txt>

- RFC 6531, "*SMTP extension for internationalized email address*", qui indique les modifications de SMTP pour utiliser le courrier aux adresses Unicode.
- RFC 6532, "*Internationalized Email Headers*", qui décrit l'utilisation d'UTF-8 dans les en-têtes du courrier électronique où, jusqu'à présent, seul ASCII était autorisé.
- RFC 6533, "*Internationalized Delivery Status and Disposition Notifications*", qui précise le format des accusés de réception et des avis de remise et de non-remise.
- Le cas des listes de diffusion est traité dans le RFC 6783.
- Et des RFC pour les protocoles permettant la récupération locale du courrier, c'est-à-dire POP (RFC 6856) et IMAP (RFC 6855), ainsi que deux RFC (RFC 6857 et RFC 6858) décrivant un algorithme pour se « replier » au cas où, à la lecture d'un message reçu, il faille transmettre un message internationalisé à un vieux logiciel ne comprenant pas la nouvelle norme.

Il existe plusieurs mises en œuvre d'EAI, dont l'interopérabilité a été testée en juillet 2008 <<http://www.ietf.org/mail-archive/web/ima/current/msg02554.html>> lors d'une session commune <<http://cslab.kr/eai-test/>> aux NIC chinois, coréens, japonais et taïwanais.

Je ne les ai pas encore déployées sur mes serveurs donc inutile d'écrire à stéphane@bortzmeyer.org, cela ne marchera pas. (Si vous voulez tester, des auto-répondeurs utilisant ces adresses sont disponibles <http://idn.icann.org/E-mail_test>.) Pire, en général, le message sera massacré par les divers intervenants (encodé en Quoted-Printable ou transformé en stphane@bortzmeyer.org ou Dieu sait quoi) **avant** même d'arriver à mon serveur. Merci à André Sintzoff pour ses tests.

L'expérience avec la syntaxe actuelle des adresses, ignorée par bien des programmeurs amateurs <<https://www.bortzmeyer.org/arreter-d-interdire-des-adresses-legales.html>>, donne à penser qu'il ne sera hélas pas possible de si tôt d'utiliser une telle adresse dans les formulaires.

Le principal protocole de la famille TCP/IP qui ne soit pas complètement internationalisé est donc désormais FTP qui dispose de deux types de transfert de données, « texte » et « binaire » où « texte » ne concerne que l'ASCII (ou, à la rigueur, les jeux ISO 8859). Un travail est en cours pour étendre FTP afin de pouvoir utiliser le texte Unicode du RFC 5198.