

Les normes du courrier électronique enfin entièrement internationalisées

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 7 Septembre 2008

<http://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-cooperation.fr` alors qu'on voudrait pouvoir écrire à `stphane@internet-en-cooprati.on.fr`. Désormais, nous avons une solution, "*Internationalized Email*", dont les normes sont désormais effectives.

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 elle avait été décrite dans le RFC 4952. Mais ce RFC ne décrivait que l'architecture générale. Il manquait les normes elle-mêmes qui, après une histoire complexe, sont désormais publiées. Pour l'instant, compte tenu de l'importance du courrier et de l'importance des modifications requises, ces spécifications ont le statut « Expérimental » et ne sont donc pas formellement sur le chemin des normes de l'IETF.

Il y a deux parties à l'adresse (décrites dans le RFC 2822) : la partie locale, à gauche du [Caractère Unicode non montré ²] 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 :

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

²Car trop difficile à faire afficher par L^AT_EX

- RFC 5335, "*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 5336, "*SMTP extension for internationalized email address*", qui indique les modifications de SMTP pour utiliser le courrier aux adresses Unicode.
- RFC 5337, "*Internationalized Delivery Status and Disposition Notifications*", qui précise le format des accusés de réception et des avis de remise et de non-remise.
- RFC 5504, "*Downgrading mechanism for Internationalized eMail Address*", qui fournit un mécanisme de **repli** ("*downgrade*") au cas où un autre serveur SMTP ne gère pas les extensions internationales,
- Et des RFC pour les protocoles permettant la récupération locale du courrier, c'est-à-dire POP (RFC 5721) et IMAP (RFC 5738).

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 taiwanais.

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.