

RFC 6530 : Overview and Framework for Internationalized Email

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 18 février 2012

Date de publication du RFC : Février 2012

<https://www.bortzmeyer.org/6530.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 standard, "*Internationalized Email Addresses*" <<https://www.bortzmeyer.org/courrier-entierement-internationalise.html>>.

Bien sûr, pour l'exemple ci-dessus, le gain n'est pas évident. Mais il l'est beaucoup plus si vous voulez écrire votre adresse avec l'écriture arabe ou chinoise. (Cf. section 3 pour la description du problème, section qui exprime clairement le fait qu'un Chinois a le droit d'avoir une adresse en sinogrammes, comme un Occidental jouit des avantages d'une adresse en caractères latins.)

Le **contenu** des courriers était internationalisé depuis longtemps, au moins depuis MIME (RFC 1341¹ à l'origine, en juin 1992). De même, certains en-têtes comme le sujet ou même comme le nom de l'expéditeur dans le champ `From:` - mais pas son adresse - pouvaient être internationalisés via MIME (RFC 2047). Mais les adresses ne l'étaient pas encore et ce manque de cohérence (possibilité d'utiliser des écritures non-latines à certains endroits mais pas à d'autres) rendait le courrier peu attrayant pour M. Ali ou M. Wu. Notre RFC, qui remplace le RFC 4952 qui n'était qu'expérimental, décrit le cadre général de l'internationalisation des adresses de courrier. Il y avait depuis longtemps des **messages internationaux**, il y a désormais également des **adresses internationales**.

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

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

de non-délivrance, par exemple, allait afficher la forme Punycode du nom. Ce n'était pas une vraie internationalisation.

Le cadre d'internationalisation du courrier couvre plusieurs aspects, couverts chacun par un RFC <https://www.bortzmeyer.org/courrier-entierement-internationalise.html>. Ce RFC expose l'architecture générale. La section 5 décrit le fonctionnement du nouveau système.

SMTP est modifié pour permettre une nouvelle extension, `SMTPUTF8`, indiquant le support du courrier internationalisé (RFC 6531).

Le format des messages (RFC 5322) est modifié pour permettre d'inclure de l'Unicode, encodé en UTF-8, directement dans les en-têtes, comme le champ `To` : (RFC 6532). Les autres "charsets" (comme UTF-32) ne sont pas prévus.

Le format des accusés de réception est modifié pour permettre des rapports de remise ou de non-remise lorsque l'expéditeur ou le destinataire avaient des adresses Unicode (RFC 6533).

Plusieurs autres protocoles « auxiliaires » seront également modifiés, par exemple pour les listes de diffusion (RFC 6783) ou pour IMAP et POP. Même chose pour des formats comme les IRI de plan `mailto` : (section 11.1 de notre RFC, RFC 3987 et RFC 6068).

Notons (section 11.2 du RFC) que les adresses de courrier sont utilisées en de nombreux endroits, par exemple comme identificateurs sur certains sites Web de commerce électronique, comme Amazon et que ces utilisateurs vont également devoir s'adapter. Quant on voit le nombre de sites, supposés professionnels, qui continuent à interdire des adresses de courrier légales <https://www.bortzmeyer.org/arreter-d-interdire-des-adresses-legales.html>, on mesure l'ampleur du travail http://www.circleid.com/posts/20091120_idn_and_email_the_harsh_reality/. La section 10, elle aussi, décrit des questions liées à l'interface utilisateur. Par exemple, elle rappelle qu'il y a toujours eu des cas délicats dans la syntaxe, dont il est plus prudent de se tenir à l'écart. Ainsi, la partie gauche d'une adresse est sensible à la casse donc, en théorie, on pourrait avoir une adresse `Stephane@bortzmeyer.fr` et que le courrier bêtement envoyé à `stephane@bortzmeyer.fr` n'atteindrait pas. Inutile de dire que peu de MTA mettent en œuvre une telle distinction... Dans le cas des adresses internationalisées, il faut donc prendre garde à la canonicalisation Unicode. Au minimum, il faudrait accepter la canonicalisation en NFC de toute adresse, car cette canonicalisation aura peut-être lieu quelque part dans le chemin, qu'on le veuille ou pas.

Les changements par rapport à l'ancien cadre (RFC 4952 et ses copains) sont décrits dans les sections 6 et 12. La section 6 fait en outre le bilan de la phase expérimentale. Désormais, les adresses de courrier internationalisées ont le statut de norme et non plus de simple expérience. Le principal changement technique, de loin, est la suppression du **repli**. Le repli ("*downgrading*", cf. RFC 5504 et RFC 5825) était prévu pour le cas où un émetteur internationalisé rencontrerait un récepteur qui ne l'est pas, nécessitant une modification des adresses en ASCII. En effet, contrairement à HTTP, le courrier ne fonctionne pas de bout en bout et le message est relayé par plusieurs serveurs, qu'on ne connaît pas forcément a priori.

Le repli était certainement la partie la plus complexe : modifier un message est toujours une opération délicate, notamment en présence de signatures cryptographiques (section 11.3), que le repli ne doit pas invalider ! Tellement complexe que sa mise en œuvre s'est avérée très délicate et que les différentes implémentations avaient du mal à interagir. En outre, l'expérience a montré que le fait de devoir gérer deux adresses (celle en Unicode et une en ASCII pour le repli), sans lien sécurisé entre les deux, était une belle source de problèmes. Résultat, par rapport aux RFC de la phase expérimentale, le repli a été abandonné. Désormais, un courrier internationalisé ne pourra tout simplement pas être transmis à

travers les vieux serveurs, ce sera au MUA ou à l'utilisateur de trouver une solution de rechange (cf. section 8, qui détaille les solutions). On note que la phase expérimentale, dont beaucoup de RFC ne sortent jamais, a été ici utile, permettant, par la pratique, d'atteindre un très large consensus sur un point qui était précédemment fortement discuté.

Un deuxième changement technique important par rapport aux premiers RFC sur ce sujet est que, en SMTP, le client aussi bien que le serveur doit signaler sa volonté d'utiliser les nouvelles adresses.

La section 8 décrit plus en détail la marche à suivre si le serveur suivant ne gère pas les adresses Unicode (par exemple, on peut prévoir que certains administrateurs activeront les adresses Unicode sur le serveur de courrier principal du domaine et oublieront le MX secondaire <<https://www.bortzmeyer.org/mx-secondaire.html>>). Elle n'est pas normative puisque c'est largement une question d'interface utilisateur, quelque chose que l'IETF ne fait pas normalement. Comme le note cette section, le meilleur repli est celui fait manuellement par l'utilisateur, car celui-ci est intelligent et motivé. Néanmoins, si on tient à faire du repli automatique, il peut être fait par le MUA, ou bien par une future extension au protocole de soumission (RFC 4409). Ça, c'était pour le cas où le serveur ancien était le premier de la chaîne. Et si le problème d'un serveur non-Unicode survient en cours de route, lorsqu'un MTA se rend compte que le MTA suivant ne convient pas? La section 9 rappelle qu'il est désormais strictement interdit aux MTA de tripoter le message et il n'y a donc rien d'autre à faire que d'indiquer une erreur, qui sera remontée à l'utilisateur.

Et si le problème survient une fois le message délivré, parce que le protocole de récupération n'a pas été adapté aux adresses Unicode? Le problème est là encore formellement en dehors du cadre, le RFC notant juste que le MDA peut faire un repli automatique s'il le juge utile.

Cette extension du courrier ne semble pas encore implémentée officiellement dans les logiciels classiques de courrier, malgré la longue période des RFC expérimentaux.