

RFC 6531 : SMTP Extension 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/6531.html>

Dans la grande série des RFC décrivant les adresses de courrier internationalisées, ce document traite des modifications au protocole SMTP. Il succède au RFC 5336¹, qui avait créé ces modifications mais qui n'avait que le statut « Expérimental ». Désormais, ces extensions internationales sont sur le Chemin des Normes de l'IETF.

Plusieurs RFC <<https://www.bortzmeyer.org/courrier-entierement-internationalise.html>> composent la nouvelle norme EAI ou « *Email Addresses Internationalized* » (cf. RFC 6530). Les deux principaux concernent le format des messages (RFC 6532) et le dialogue entre deux serveurs de messagerie, suivant le protocole SMTP. Ce dernier cas est couvert dans notre RFC 6531.

Les adresses de courrier en Unicode ne sont pas en effet compatibles avec le SMTP « brut ». Ainsi, la commande RCPT TO qui permet d'indiquer le destinataire, ne prend en paramètre que de l'ASCII. Et les en-têtes du courrier peuvent désormais être en UTF-8 (RFC 6532). Il faut donc étendre SMTP (actuellement normalisé en RFC 5321) pour permettre de l'UTF-8 en paramètre de commandes comme MAIL FROM et RCPT TO. Cette extension se signale avec l'option SMTPUTF8 (l'ancien RFC, le RFC 5336, utilisait UTF8SMTP), dont l'usage permet de vérifier que le MTA situé en face comprend bien la nouvelle norme (sections 1 et 2 du RFC).

La section 3 forme le gros du RFC en spécifiant rigoureusement le nouveau protocole. Voyons d'abord un exemple de dialogue SMTP avec la nouvelle extension (C est le client - l'appelant - et S le serveur - l'appelé). Notre RFC 6531 ne modifie pas SMTP (RFC 5321), il l'étend :

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc5336.txt>

```

S: 220 mail.example.org ESMTP FooBar (Debian/GNU)
C: EHLO mail.iloveunicode.example
S: 250-mail.example.org
  250-ENHANCEDSTATUSCODES
  250-8BITMIME
  250-SMTPUTF8
  250 DSN
C: MAIL FROM:<stéphane@internet-en-coopération.fr> SMTPUTF8
S: 250 2.1.0 Ok
C: RCPT TO: <françoise@comptabilité.example.org>
S: 250 2.1.5 Ok

```

Notez les deux adresses, chacune comprenant des caractères non-ASCII, à gauche et à droite du @.

Les serveurs qui mettent en œuvre l'extension SMTPUTF8 doivent également accepter la très ancienne extension 8BITMIME (RFC 6152) qui permet, depuis belle lurette, de ne pas se limiter à l'ASCII dans le contenu des messages, sans pour autant nécessiter des encodages comme Quoted-Printable (en dépit de ce que prétend une légende tenace).

Les sections 3.1 et 3.2 présentent la nouvelle extension, SMTPUTF8 (enregistrée à l'IANA <<https://www.iana.org/assignments/mail-parameters>>, cf. section 4.1). Les extensions SMTP sont définies par le RFC 5321, section 2.2. Un MTA peut annoncer qu'il accepte SMTPUTF8, permettant à son pair, s'il l'accepte également, d'utiliser des adresses UTF-8. C'est également cette section qui définit ce que doit faire un MTA lorsqu'il tente de transmettre un message internationalisé à un ancien MTA qui ne gère pas cette extension. Trois solutions sont possibles, dont le rejet complet du message (retour à l'expéditeur) et un effort pour trouver une route alternative, passant par des serveurs plus modernes (le RFC 5336 avait un quatrième mécanisme, le repli - "downgrade" - qui a été abandonné après usage, car trop complexe et trop imprédictible).

La section 3.3 décrit la nouvelle syntaxe pour les adresses. L'ancienne était limitée à ASCII mais très permissive (bien qu'on trouve, notamment derrière les formulaires Web, énormément de logiciels de « vérification » bogués <<https://www.bortzmeyer.org/arreter-d-interdire-des-adresses-legales.html>>). En gros, la syntaxe est qu'une adresse est formée de deux parties, la partie locale et le nom de domaine, tous les deux séparés par l'arobase. La nouveauté est que les deux parties peuvent désormais être en Unicode, avec peu de restrictions sur la partie locale (le nom de domaine étant soumis aux restrictions du RFC 5890).

Il y a plein d'autres détails dans cette section 3, comme celui (section 3.7.1) que le paramètre de la commande EHLO (un nom de machine) doit être encodé en Punycode puisque, à ce stade, on ne sait pas encore si SMTPUTF8 est accepté ou pas.

Les changements depuis le RFC 5336 sont résumés dans la section 6 du RFC 6530. Le principal est la suppression de la possibilité de **repli** ("downgrading") automatique vers des adresses en ASCII pur. Le paramètre ALT-ADDRESS disparaît donc.

Cette extension de SMTP ne semble pas mise en œuvre dans les grands logiciels libres classiques comme sendmail ou exim. Pour Postfix, l'auteur a clairement exprimé son manque d'enthousiasme <http://www.engardelinux.org/modules/index/list_archives.cgi?list=postfix-devel&page=0001.html&month=2011-10>. Il existe des "patches" non-officiels développés en Chine, pour Postfix <<http://www.cdnc.org/gb/news/src-EAI.tar.gz>> (l'archive comprend également d'autres logiciels, d'où sa taille) et pour sendmail <<http://www.cdnc.org/gb/news/twnic.zip>>. La société chinoise Coremail est en train de travailler à un produit commercial basé sur ces "patches".