

RFC 5336 : SMTP extension for internationalized email address

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 7 Septembre 2008

Date de publication du RFC : Septembre 2008

<http://www.bortzmeyer.org/5336.html>

Dans la grande série des RFC décrivant les adresses de courrier internationalisées, ce document traite des modifications au protocole SMTP.

Plusieurs RFC composent la nouvelle norme EAI ou « *Email Addresses Internationalized* » (cf. RFC 4952¹). Les deux principaux concernent le format des messages (RFC 5335) et le dialogue entre deux serveurs de messagerie, suivant le protocole SMTP. Ce dernier cas est couvert dans notre RFC 5336.

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. 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 `UTF8SMTP`, dont l'usage permet de vérifier que le MTA situé en face comprend bien la nouvelle norme (sections 1 et 2 du RFC). Si ce n'est pas le cas, une adresse en ASCII facultative peut être utilisée.

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) :

1. Pour voir le RFC de numéro NNN, <http://www.ietf.org/rfc/rfcNNN.txt>, par exemple <http://www.ietf.org/rfc/rfc4952.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-UTF8SMTP
  250 DSN
C: MAIL FROM:<stéphane@internet-en-coopération.fr> ALT-ADDRESS=bortzmeyer@afnic.fr
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 UTF8SMTP doivent également accepter la très ancienne extension 8BITMIME (RFC 1652) 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 d'une légende tenace).

La section 3.1 présente la nouvelle extension, UTF8SMTP. Les extensions SMTP sont définies par le RFC 5321, section 2.2. Un MTA peut annoncer qu'il accepte UTF8SMTP, 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. Quatre solutions sont possibles, la plus fréquente sera sans doute d'utiliser le mécanisme de repli ("*downgrade*").

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 <<http://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 3490).

Une adresse alternative tout en ASCII peut être fournie par le paramètre ALT-ADDRESS. C'est l'objet des sections 3.4 et 3.5. Elle doit s'utiliser conjointement à une modification du message (qui peut contenir des en-têtes en UTF-8) et l'ensemble du processus s'appelle le **repli** ("*downgrade*"). Sa spécification est publiée dans le RFC 5504. (Dans l'exemple de dialogue SMTP plus haut, seul l'expéditeur avait une adresse alternative.)

Il y a plein d'autres détails dans cette section, 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 UTF8SMTP est accepté ou pas.

Je n'ai pas encore testé avec mes Postfix, ni vérifié quels étaient les plans des auteurs de ce logiciel vis-à-vis de cette extension de SMTP.