

RFC 5536 : Netnews Article Format

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 29 novembre 2009

Date de publication du RFC : Novembre 2009

<https://www.bortzmeyer.org/5536.html>

Malgré la concurrence de la messagerie instantanée, du courrier, du Web avec la syndication, Usenet / Netnews, un des plus anciens outils numériques de communication et de distribution d'information continue à vivre. Plus ancien que l'Internet sur beaucoup de sites (Usenet était souvent distribué via UUCP), ce dinosaure continue à faire la joie et le désespoir de millions d'utilisateurs. Mais ses normes avaient un peu vieilli et pas subi de mise à jour depuis longtemps. C'est heureusement ce qui vient de leur arriver, dans une série de RFC dont les premiers publiés sont le RFC 5537¹ et notre RFC 5536, qui normalise le **format** des messages. Ces documents ont pris plus de huit ans à être élaborés et approuvés.

Notre RFC 5536 est le successeur du célèbre RFC 1036, publié en décembre 1987, et qui spécifiait avant lui le format des messages (comme ce que fait le RFC 5322 pour le courrier). Depuis des années, les mises en œuvres existantes de Usenet utilisaient un sur-ensemble du RFC 1036, sur-ensemble qui vient d'être normalisé. (Avant cela, ces améliorations étaient contenues dans un document informel, « *Son of RFC 1036* », qui a finalement été publié en RFC 1849.)

Les "news" comprennent plusieurs parties : le protocole de transport des messages, aujourd'hui essentiellement NNTP (RFC 3977), le format des messages (notre RFC 5536) et des conventions plus ou moins formelles entre les sites qui s'envoient des messages NetNews, ces sites formant le réseau Usenet (ou des réseaux privés).

Que contient concrètement notre RFC ? La section résume les concepts de base. « NetNews » est composé d'un ensemble de protocoles pour stocker, transmettre et lire des **articles**, organisés en **groupes** ("newsgroups"). Ces articles sont typiquement distribués au sein d'Usenet par un algorithme d'inondation. Pour retrouver facilement un article, et déterminer s'il est déjà présent localement, le serveur Usenet

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc5537.txt>

utilise un "Message ID" (section 3.1.3) qui est présent dans chaque article, et est unique. Le format NetNews est très proche de celui du courrier, tel que décrit dans les RFC 5322 et RFC 2045. La définition de la syntaxe d'un article (section 1.4) emprunte également beaucoup au RFC 5322 (voir annexe C, qui développe les différences entre les deux formats; notre RFC 5536 a une syntaxe plus stricte).

La section 2 décrit en détail ce format, en partant du RFC 5322, mais en ajoutant des restrictions (section 2.2). L'internationalisation, comme pour le courrier, est assurée par MIME (section 2.3) et notamment les RFC 2049 et RFC 2231 (voir aussi la section 4).

Une fois les principes de base du format posés dans la section 2, la section suivante décrit chaque en-tête possible. Certains sont obligatoires dans tout article (section 3.1) comme `Date:` (section 3.1.1) ou `From:` (3.1.2). Est également obligatoire `Message-ID:` (section 3.1.3) qui est un identifiant unique de l'article, sur lequel les serveurs se basent pour mettre en œuvre l'algorithme d'inondation. Les identificateurs d'articles sont transmis à tous les serveurs voisins, ceux-ci pouvant alors dire s'ils acceptent ou refusent (car ils ont déjà un article de même `Message-ID:`). (Le `Message-ID:` sert également aux recherches d'un article, par exemple, via GoogleGroups <http://groups.google.com/advanced_search>.)

Dans les en-têtes obligatoires, on trouve également `Newsgroups:` (section 3.1.4) qui donne la liste des groupes dans lequel l'article a été envoyé et `Path:` (3.1.5), qui est la séquence des serveurs par lesquels est passé l'article, séparés par des!. Cet en-tête vise à empêcher les boucles dans l'algorithme d'inondation : même si la détection d'un `Message-ID:` connu ne marche pas, le fait que le serveur apercevra son nom dans le `path:` lui indiquera qu'il doit rejeter l'article.

L'ancienneté des NetNews fait que beaucoup de points de la norme s'expliquent par des raisons historiques. Ainsi, le fait de mettre le nom de l'expéditeur au début du `Path:` et la ressemblance d'un `Path:` avec une adresse UUCP avec route explicite ne sont pas un hasard. Dans les réseaux UUCP, le `Path:` pouvait être utilisé comme adresse de courrier, pour écrire à l'expéditeur (cet usage s'est perdu et notre RFC recommande de mettre `not-for-mail` comme adresse d'expéditeur).

Bien sûr, il y a aussi des en-têtes facultatifs, décrits dans la section 3.2 : c'est le cas par exemple de `Archive:` (section 3.2.2) qui indique si l'expéditeur veut que son message puisse être archivé (Google permet de chercher dans tous les articles échangés sur Usenet depuis 1981 <<http://groups.google.com/>>).

`Expires:` (section 3.2.5) sert à indiquer au bout de combien de temps l'article peut être supprimé d'un serveur qui veut faire de la place (en l'absence de cet en-tête, le serveur décide seul de la date d'expiration).

Et il y a bien d'autres en-têtes possibles dans cette section 3.2.

Quelle est la sécurité offerte par Usenet? Nulle, rappelle la section 5. Les articles ne sont pas confidentiels et leur intégrité n'est pas garantie.

La liste complète des changements depuis le RFC 1036 apparaît en annexe B. MIME est désormais formellement reconnu (la plupart des logiciels le géraient depuis longtemps mais il n'existait pas à l'époque du RFC 1036). Certains en-têtes largement acceptés par les logiciels deviennent officiels, comme `Archive:`. Autrement, les changements sont surtout du toilettage de syntaxe.

Voici, tel qu'il passe sur le réseau (affiché dans un logiciel de lecture, cela peut évidemment être très différent) un article à ce format :

Path: news.free.fr!xref-2.proxad.net!spooler1c-1.proxad.net!cleanfeed2-a.proxad.net!nnrp3-2.free.fr!not-for-mail
From: Stephane Acounis <panews@free.fr>
Subject: =?iso-8859-1?b?wA==? donner: Sun(s) SS20 sur Nantes
Date: Mon, 09 Jun 2008 19:24:17 +0200
User-Agent: Pan/0.14.2 (This is not a psychotic episode. It's a cleansing moment of clarity.)
Message-Id: <pan.2008.06.09.17.24.16.590396@free.fr>
Newsgroups: fr.comp.ordinosaures
MIME-Version: 1.0
Content-Type: text/plain; charset=ISO-8859-15
Content-Transfer-Encoding: 8bit

Bonsoir,

J'ai deux stations Sun SS20, une bi-pro (SM75), l'autre mono-pro (SM75),
à donner sur Nantes (ou Nançay mercredi/jeudi ou Brest vendredi).
256Mo de mémoire vive, disque 9Go, carte vidéo.
Je doit avoir des claviers, des câbles divers, des cartes SBus en plus.

...

Notez l'en-tête Subject : encodé selon le RFC 2047.

NetNews est mis en œuvre dans de nombreux logiciels aujourd'hui, souvent dans les logiciels de courrier, vu la ressemblance des formats. C'est ainsi que le célèbre MUA Thunderbird est également un bon lecteur de News.