Les spammeurs ne sont même pas compétents en standards du courrier

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 20 janvier 2018

http://www.bortzmeyer.org/mime-newsletter.html

Comme vous, je reçois du spam via le courrier électronique. J'en reçois peut-être davantage car mon adresse publiée sur ce blog a manifestement été mise sur des listes d'« influenceurs » et autres blogueurs mode. Ces spams, plus orientés « professionnels » sont souvent en deux parties, une en texte seul et l'autre en HTML. Mais ce qui est amusant est que les deux parties sont parfois désynchronisées.

[Si vous connaissez bien MIME, vous pouvez sauter l'essentiel de cet article, et aller directement aux observations de la fin. J'explique d'abord MIME pour les gens qui ne connaissent pas.] Cette possibilité d'envoyer un message en plusieurs parties est normalisée dans le standard dit MIME, spécifié dans le RFC 2045[1]. Il existe de nombreuses possibilités de message en plusieurs parties dans MIME dont les deux plus connues sont les parties **différentes** (`multipart/mixed` dans les étiquettes MIME) et les parties **alternatives** (`multipart/alternative` pour MIME). Bien sûr, la plupart des gens qui envoient des "*newsletter*" et CP ne savent pas cela. Ils ignorent que les messages ont une forme normalisée dans le RFC 5322 et qui est assez différente de ce qu'ils voient sur leur écran. Ainsi, prenons un message qui ressemble à ceci avec Thunderbird :

Sur le réseau, il sera très différent. Ce qui passe, et qui est interprété par Thunderbird, sera :

```
Date: Sat, 20 Jan 2018 17:07:29 +0100
From: Stephane Bortzmeyer <stephane@bortzmeyer.org>
To: johndoe@example.com
Subject: Les roux, c'est sympa
Message-ID: <20180120160727.ehuvwpqhrhbzpv26@bortzmeyer.org>
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary="gr5ck6jng3husyqz"
User-Agent: NeoMutt/20170113 (1.7.2)
```

1. Pour voir le RFC de numéro NNN, `https://www.ietf.org/rfc/rfcNNN.txt`, par exemple `https://www.ietf.org/rfc/rfc2045.txt`

```
--gr5ck6jng3husyqz
Content-Type: text/plain; charset=utf-8
Content-Disposition: inline
Content-Transfer-Encoding: 8bit

Un potamochère roux au ZooParc de Beauval.
Date        15 May 2016, 16:03:54
Source        Own work
Author        Thesupermat

--gr5ck6jng3husyqz
Content-Type: image/jpeg
Content-Disposition: attachment; filename*=utf-8''Zooparc_de_Beauval_-_Potamoch%C3%A8re_roux_-_2016_-_002%2
Content-Transfer-Encoding: base64

/9j/4AAQSkZJRgABAgEBLAEsAAD/4UnURXhpZgAASUkqAAgAAAAMAA8BAgAGAAAAngAAABAB
AgAVAAAApAAAABIBAwABAAAAAQAAABoBBQABAAAAugAAABsBBQABAAAAwgAAACgBAwABAAAA
AgAAADEBAgAVAAAAygAAADIBAgAUAAAA4AAAADsBAgAMAAAA9AAAAJiCAgAJAAAAAAEAAGmH
...
```

Les parties différentes sont utilisées pour le cas où on envoie, par exemple, un texte et une image d'illustration (ce qui est le cas de l'exemple ci-dessus). Elles sont décrites dans le RFC 2046, section 5.1.3. Le message a un en-tête du genre (regardez l'exemple plus haut) :

```
Content-Type: multipart/mixed; boundary="------------514DC64B3521BD0334B199AA"
```

Une autre possibilité courante de message en plusieurs parties est le cas des parties alternatives (RFC 2046, section 5.1.4). Prenons ce spam typique :

Sur le réseau, le contenu avait en fait deux parties, une en texte brut et une en HTML. Normalement, les deux sont équivalentes, et le MUA va choisir laquelle afficher. Dans l'image ci-dessus, Thunderbird affichait le HTML. Ici, un autre MUA, mutt fait un autre choix et affiche le texte seul :

Quant à ce qui circulait sur le câble, cela ressemble à :

```
MIME-Version: 1.0
To: bortzmeyer+ietf@nic.fr
Content-Type: multipart/alternative;
boundary="----=_NextPart_001_322D_019C6BD8.657D49B2"
X-Mailer: Smart_Send_2_0_138
Date: Thu, 18 Jan 2018 08:49:31 -0500
Message-ID: <586044135608831506631O@Ankur>
X-SMTPCOM-Tracking-Number: f5f27516-49b6-40ee-848b-62cb2597a453
X-SMTPCOM-Sender-ID: 6008902
Feedback-ID: 6008902:SMTPCOM
Subject: The SDN, NFV & Network Virtualization Ecosystem: 2017 - 2030 - Opportunities, Challenges, Strategie
)
From: "Andy Silva" <andy.silva@snsreports.com>


------=_NextPart_001_322D_019C6BD8.657D49B2
Content-Type: text/plain; charset="iso-8859-1"
Content-Transfer-Encoding: quoted-printable

The SDN, NFV & Network Virtualization Ecosystem: 2017 =96 2030 =96 Opportun=
```

```
ities, Challenges, Strategies & Forecasts

...

------=_NextPart_001_322D_019C6BD8.657D49B2
Content-Type: text/html; charset="iso-8859-1"
Content-Transfer-Encoding: quoted-printable

<HTML xmlns:o><HEAD>
<META content=3D"text/html; charset=3Diso-8859-1" http-equiv=3DContent-Type>
<META name=3DGENERATOR content=3D"MSHTML 11.00.10570.1001"></HEAD>
<BODY>
<P><SPAN><FONT color=3D#000000><FONT color=3D#000000><FONT color=3D#000000>=
...
```

Enfin, il existe d'autres types de messages en plusieurs parties, comme les possibilités multilingues du RFC 8255.

Voilà, tout cela est ancien (le RFC 2046 date de 1996 !) Maintenant, passons aux observations récentes. J'ai reçu le 15 janvier 2018 un spam de l'agence Maatch annonçant un « Voyage de presse Roanne Tout & Simplement » (un voyage de presse est un événement où une enterprise nourrit et abreuve des « influenceurs » pour qu'ils écrivent ensuite des articles favorables). Bizarrement, le voyage est annoncé pour le 9 septembre 2016. Et le message est « signé » d'un nom différent de celui utilisé dans le champ `From:`. Il se trouve que je lis mon courrier avec mutt et l'option `alternative_order text/plain` qui lui indique de préférer, dans les messages composés de plusieurs parties alternatives, le texte seul. Si je regarde la partie HTML du message, je vois la bonne date (31 janvier 2018), le bon programme (« invitation au restaurant gastronomique Le Central, en compagnie de l'équipe de Roanne Tout & Simplement et de chefs d'entreprises numériques ») et la bonne « signature ». La personne qui a préparé le message n'a sans doute pas directement écrit au format IMF ("*Internet Message Format*", RFC 5322) mais a utilisé un logiciel qui n'affichait que la partie HTML, la seule qu'elle met à jour quand elle prépare un nouvel envoi. Toutes les occurrences futures de ce voyage de presse seront donc envoyés avec une partie en texte brut restée figée en 2016. Et personne ne s'en est aperçu car l'agence ne teste pas ses messages avec divers MUA et que les destinataires, soit sont mariés avec les logiciels « HTML seulement », soit envoient automatiquement ce genre de messages dans le dossier Spam.

Le choix de l'utilisateur, présenté par mutt :

Est-ce un cas isolé dû à une agence de communication particulièrement incompétente ? Non, j'ai regardé d'autres spams et le phénomène ne semble pas exceptionnel. Par exemple, le 18 janvier, je reçois un spam de l'agence Srvsaas/Sellinity intitulé « Invitation personnelle : Dîner privilégié » et c'est encore plus amusant. La partie texte décrit un diner avec les commerciaux d'un fabricant de matériel réseau bien connu, la partie HTML un diner avec les commerciaux d'une **autre** entreprise, une entreprise de sécurité informatique importante. L'agence a repris un message conçu pour un client, a modifié la partie HTML pour un autre client et a oublié la partie texte !

D'autres accidents sont possibles avec ces deux parties pas toujours synchronisées. Par exemple, on me signale que les messages de confirmation de Ryanair ont une partie texte et une partie HTML et que la partie texte contient. . .du HTML. Bref, regarder la partie texte **et** la partie HTML des spams peut être une distraction intéressante.