RFC 6122 : Extensible Messaging and Presence Protocol (XMPP): Address Format

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 31 mars 2011

Date de publication du RFC : Mars 2011

http://www.bortzmeyer.org/6122.html

Ce RFC normalisait le format des **adresses** du protocole de messagerie instantanée XMPP, également connu sous son ancien nom de Jabber. Il s'agissait d'un document transitoire, qui remplaçait une partie de l'ancien RFC 3920 [1], et qui a lui même été remplacé par le RFC 7622.

Le protocole XMPP est le standard actuel de messagerie instantanée de l'IETF. Les utilisateurs sont identifiés par une **adresse** également connue, pour des raisons historiques, sous le nom de **JID** (*"Jabber IDentifier"*). Ces adresses ont une forme qui ressemble aux adresses de courrier électronique mais elles sont en fait complètement différentes. Ainsi, j'ai personnellement deux adresses XMPP, bortzmeyer@gmail.com (le service de messagerie Google Talk utilise en effet la norme XMPP), qui me sert surtout pour la distraction, et bortzmeyer@dns-oarc.net qui me sert pour le travail. Mais rien ne dit qu'un courrier envoyé à ces adresses fonctionnera : adresses XMPP et de courrier vivent dans des mondes différents.

Le format des adresses XMPP était spécifié à l'origine dans le document XEP-0029 <http://xmpp.org/extensions/xep-0029.html>. Puis il a été normalisé dans la section 3 du RFC 3920. La norme XMPP ayant subi une refonte complète, qui a mené entre autre à un nouveau RFC, le RFC 6120, se posait la question du format d'adresses, dont l'internationalisation, compte-tenu de la nouvelle norme IDN <http://www.bortzmeyer.org/idnabis.html>, suscitait des débats. Ces débats n'étant pas terminés, la décision a été prise de sortir le format d'adresses du RFC principal, et d'en faire ce RFC 6122, jusqu'à ce qu'un consensus apparaisse sur le format « définitif ».

Donc, quels sont les points essentiels des adresses XMPP (section 2) ? Une adresse (ou JID) identifie une entité (un humain mais peut-être aussi un programme automatique). Elle comprend trois parties (donc une seule est obligatoire), la **partie locale**, le **domaine** et une **ressource**. Ainsi, dans bortzmeyer@gmail.com/Home,

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc3920.txt

`bortzmeyer` est la partie locale, `gmail.com` le domaine et `Home` la ressource. Le @ et le / servent de séparateurs. Chacune des trois parties peut-être composée de caractères Unicode (cf. section 3), qui doivent être encodés en UTF-8. La partie locale doit être canonicalisée avec `Nodeprep`, le domaine avec le `Nameprep` du RFC 3491. Seul le domaine est obligatoire et `bot.example.org` est donc un JID valide.

Les adresses XMPP (ou JID) sont souvent représentées sous forme d'un IRI, selon le RFC 5122. En gros, un IRI XMPP est un JID préfixé par `xmpp:`. Mais ces IRI ne sont pas utilisés par le protocole XMPP. Dans les champs `to` et `from` des strophes ("*stanzas*") XMPP, on ne trouve que des JID (sans le `xmpp:` devant). Les IRI ne servent que dans les liens dans les pages Web, comme `<xmpp:bortzmeyer@gmail.com>` (attention, ce lien ne marchera pas avec tous les navigateurs).

Le RFC détaille ensuite chaque partie. Le domaine, seule partie obligatoire, est dans la section 2.2. On peut le voir comme une identification du service auquel on se connecte (`gmail.com` = Google Talk), et la création de comptes et l'authentification se font typiquement en fonction de ce service. En théorie, une adresse IP est acceptable pour cette partie mais, en pratique, c'est toujours un FQDN. Ce nom de domaine peut être un IDN donc `instantanée.nœud.example` est un nom acceptable pour former un JID. Auquel cas, ce doit être un IDN légal (ce qui veut dire que toute chaîne de caractères Unicode n'est pas forcément un nom légal pour la partie domaine d'un JID). À noter que XMPP n'utilise **pas** l'encodage en ASCII des IDN (le Punycode).

Et la partie locale, celle avant le @ ? La section 2.3 la couvre. Elle est optionnelle et identifie en général une personne (mais peut aussi indiquer un programme ou bien un salon de conversation à plusieurs). Si elle est en Unicode, elle doit être canonicalisable avec le profil `Nodeprep` (`Nodeprep` est normalisé en annexe A de notre RFC).

Quant à la ressource, troisième partie du JID (après le /), également optionnelle, elle sert à distinguer plusieurs sessions par le même utilisateur (par exemple `/Home` et `/Office`). La section 2.4 la décrit en détail. Également en Unicode, elle est canonicalisée par le profil `Resourceprep`. Elle n'a pas de structure officielle. Un / dans une ressource (par exemple `example.com/foo/bar`) n'implique donc pas une hiérarchie. De même, si on y trouve quelque chose ressemblant à une adresse (par exemple `joe@example.net/nic@host`), cette ressource `nic@host` doit être traitée comme un identificateur opaque (et pas être analysée en « nom (at) machine »).

Quelles sont les conséquences de sécurité de ces adresses ? Il n'y en a guère mais, bon, il y a toujours des inquiets donc la section 4 examine en détail tous les risques, même très théoriques. Bien sûr, les adresses XMPP héritent des questions de sécurité de `Stringprep` (section 9 du RFC 3454) mais le RFC mentionne surtout les risques de triche sur les adresses. Il y en a de deux sortes, les usurpations et les imitations. Les usurpations (section 4.3.1) sont les cas où un méchant arrive à envoyer un message XMPP en trichant sur l'origine, par exemple si `jean@example.net`, une fois authentifié auprès de son propre serveur, arrive à transmettre un message prétendant venir de `jeanne@jabber.example`. Normalement, le protocole XMPP empêche cela, par l'authentification mutuelle des serveurs (sauf attaques un peu sophistiquées, par exemple sur le DNS). Cela dit, cette authentification n'empêche pas un serveur d'annoncer une autre partie locale (`jeanne@example.net` au lieu `jean@example.net`).

L'autre risque est celui d'imitation (section 4.3.2). Cette fois, il s'agit d'utiliser des JID légitimes et authentiques mais qui ressemblent à celui de la victime, par exemple `fric@paypa1.com` au lieu de `fric@paypal.com` (si vous ne voyez pas la différence, regardez mieux). Cette technique est souvent connue sous le nom de "*typejacking*". Dans un accès de beaufitude, le RFC mentionne aussi le cas d'écritures « moins familières » que l'alphabet latin (moins familières pour qui ?) mais, comme le montre l'exemple du RFC ou bien celui cité plus haut, le problème arrive même en se limitant à ASCII. Si vous

voulez un joli exemple avec Unicode (mais le résultat dépend de l'environnement avec lequel vous lisez cet article), le RFC cite `[Caractère Unicode non montré [2] ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ]` qui ressemble à `STPETER` mais est écrit en Cherokee. Comme un JID peut contenir à peu près n'importe quel caractère Unicode, il n'y a pas vraiment de prévention technique possible contre ce problème. Toutefois, contrairement à ce que prétend le RFC (qui évoque le spectre du hameçonnage), il est peu probable que cela ait des conséquences en pratique <http://www.bortzmeyer.org/idn-et-phishing.html>. Le RFC, qui exagère le risque, suggère que les serveurs XMPP qui servent une communauté homogène, linguistiquement parlant, peuvent réduire le jeu de caractères Unicode acceptés comme partie locale d'un JID (par exemple, à l'alphabet cyrillique pour un serveur russe). Une telle politique serait, je trouve, un gros recul dans l'internationalisation de l'Internet et, de toute façon, elle n'aiderait pas les serveurs internationaux comme `jabber.org` ou Google Talk. Une approche moins restrictive est aussi proposée par le RFC : le logiciel client XMPP pourrait prévenir visuellement l'utilisateur lorsqu'un JID utilise une écriture distincte de celle utilisée habituellement par l'utilisateur (et qui est déduisible de ses préférences).

Le RFC se termine par les annexes normalisant les profils `Nodeprep` et `Resourceprep`. Rappelons que `Stringprep` (RFC 3454) décrit un cadre général pour la canonicalisation de chaînes de caractères Unicode, afin de faciliter les comparaisons ultérieures. `Stringprep` laisse un grand nombre de choix ouverts et doit donc, avant utilisation, être réduit à un profil. Les profils normalisés sont enregistrés à l'IANA <https://www.iana.org/assignments/stringprep-profiles/stringprep-profiles.xml>. Par exemple, `Nodeprep` précise que la normalisation Unicode utilisée doit être la NFKC et donne la liste des caractères interdits dans une partie locale (par exemple @, ¡, ¿, /, etc). `Resourceprep`, lui, est dans l'annexe B et n'est pas équivalent (par exemple, les caractères cités plus haut ne sont **pas** interdits dans la ressource d'un JID). Stringprep ayant depuis été abandonné (remplacé par le cadre du RFC 7564), une nouvelle version de XMPP est en cours de rédaction pour la partie « Unicode ».

La liste complète des changements par rapport au RFC 3920 figure en annexe C. Rappelons que le RFC 3920 s'appuyait, pour les parties en Unicode, sur des profils `Stringprep` (RFC 3454) comme `Nameprep`. Ce `Nameprep`, décrit dans le RFC 3491, ayant été supprimé à l'occasion de la réforme IDNA bis <http://www.bortzmeyer.org/idnabis.html>, la question se posait de l'avenir des profils `Stringprep`. Un groupe de travail IETF, PRECIS <http://tools.ietf.org/wg/precis>, a été monté pour discuter de l'avenir des normes utilisant `Stringprep`. Son travail n'étant pas terminé, et le groupe de travail sur XMPP ne souhaitant pas retarder la sortie de la nouvelle norme (RFC 6120), la question du successeur de `Nameprep` a été remise à plus tard et notre RFC 6122 normalise donc le format d'adresses XMPP tel qu'il était avant IDNA bis. Le RFC note que les implémenteurs sont encouragés à utiliser dès maintenant IDNA bis, même avant la sortie du RFC sur le format d'adresses (à part la question de la canonicalisation avec `Nameprep`, les différences entre l'IDN original et IDNA bis sont de peu d'importance pratique). Ce RFC a finalement été le RFC 7622, qui est désormais la norme des adresses XMPP.

Les autres changements sont limités : clarification sur l'utilisation d'adresses IPv6 littérales dans les JID (j'offre une bière au premier qui annonce un JID `quelqun@[2001:db8::bad:dcaf:1]` **qui marche**), correction de bogues dans la grammaire pour le cas de termes de longueur nulle, changements de terminologie (la partie locale se nommait "*node identifier*", d'où le terme de `Nodeprep`, survivance de l'ancien vocabulaire), etc.

2. Car trop difficile à faire afficher par LaTeX