RFC 8222 : Selecting Labels for Use with Conventional DNS and Other Resolution Systems in DNS-Based Service Discovery

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 22 septembre 2017

Date de publication du RFC : Septembre 2017

https://www.bortzmeyer.org/8222.html

DNS-SD ("*DNS-based Service Discovery*", normalisé dans le RFC 6763 [1]) permet de découvrir des services dans le réseau, via le DNS, **mais aussi** via d'autres systèmes de résolution de noms, comme mDNS (RFC 6762). Ces différents systèmes de résolution ayant des propriétés différentes, il peut se poser des problèmes d'interopérabilité. Ce RFC documente ces problèmes et est donc une lecture recommandée pour les développeurs DNS-SD.

Rappelons un peu de contexte (section 1 du RFC) : les applications qui utilisent le DNS imposaient fréquemment une syntaxe réduite aux noms manipulés, la syntaxe « LDH » ("*[ASCII] Letters, Digits and Hyphen*"), décrite dans le RFC 952. (Contrairement à ce qu'on lit souvent, le DNS n'impose **pas** une telle limite, elle est purement dans les applications <https://www.bortzmeyer.org/host-vs-domain.html>.) Du fait de ces règles des applications <https://www.bortzmeyer.org/pourquoi-idn-et-pas-un-dns-unicode.html>, il a fallu, pour utiliser des lettres Unicode, un système spécial, IDN (RFC 5890). IDN ajoute ses propres contraintes, comme le fait que les « non-lettres » (par exemple les emojis) ne sont pas acceptés.

Le DNS accepte tout à fait du binaire quelconque dans les noms (donc, par exemple, des caractères non-ASCII avec des encodages comme Latin-1). mDNS (RFC 6762) autorise également les caractères non-ASCII, mais impose l'encodage UTF-8. mDNS autorise des caractères qui ne sont pas permis en IDN, du banal espace aux symboles les plus rigolos. DNS-SD recommande même leur usage (RFC 6763, section 4.1.3). La seule contrainte est de se limiter au format Unicode du réseau, décrit dans le RFC 5198.

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc6763.txt

Bien des développeurs d'application ne sont pas au courant de ces subtilités. Ils manipulent des noms de services comme des chaînes de caractères, et ne se soucient pas du système de résolution sous-jacent et de ses particularités. Or, potentiellement, un même nom peut être accepté par certains de ces systèmes de résolution et refusés par d'autres.

Si le développeur est au courant, et s'il prend soin de faire en sorte que les noms « avancés » ne soient pas envoyés aux systèmes de résolution les plus anciens, ça peut marcher. Mais il ne faut pas se faire d'illusion, les fuites se produisent et tout nom utilisé sur une machine connectée à l'Internet finira tôt ou tard par arriver au DNS, comme le montre le trafic que reçoivent les serveurs de noms de la racine.

Notez enfin que les utilisateurs de DNS-SD, encore plus que ceux des noms classiques, ne tapent que très rarement les noms : presque toujours, ils les choisissent dans une liste.

Maintenant, pourquoi y a-t-il un problème ? (Section 2 du RFC.) Comme je l'ai indiqué plus haut, le DNS accepte n'importe quoi dans un nom. Pourquoi ne pas juste dire « envoyons le nom "Joe's printer, first floor" directement dans le DNS, et ça marchera » ? Il y a deux problèmes avec cette approche. D'abord, l'application qui ne ferait pas attention et enverrait un nom non-ASCII en disant « le DNS s'en tirera de toute façon », cette application oublie qu'il peut y avoir sur le trajet de sa requête une couche logicielle qui fait de l'IDN et va donc encoder en Punycode (transformation du "*U-label*" en "*A-label*"). Le « vrai » système de résolution ne verra donc pas le nom original. Un problème du même genre arrive avec certains logiciels qui se mêlent de politique, par exemple les navigateurs Web comme Firefox qui se permettent d'afficher certains noms en Unicode et d'autres en ASCII, selon la politique du TLD. Bref, le trajet d'un nom dans les différents logiciels est parsemé d'embûches si on est en Unicode.

En outre, l'approche « j'envoie de l'UTF-8 sans me poser de questions » (suggérée par la section 4.1.3 du RFC 6763) se heurte au fait que la racine du DNS et la plupart des TLD ne permettent de toute façon par d'enregistrer directement de l'UTF-8. (Au passage, le RFC oublie un autre problème de l'UTF-8 envoyé directement : les serveurs DNS ne font pas de normalisation Unicode, et l'insensibilité à la casse du DNS n'est pas évidente à traduire en Unicode.)

La section 3 de notre RFC propose donc de travailler à un profil permettant une meilleure interopérabilité. Un profil est une **restriction** de ce qui est normalement permis (un dénominateur commun), afin de passer à peu près partout. Un exemple de restriction serait l'interdiction des majuscules. En effet, le DNS est insensible à la casse mais IDN interdit les majuscules (RFC 5894, sections 3.1.3 et 4.2) pour éviter le problème de conversion majuscules¡-¿minuscules, qui est bien plus difficile en Unicode qu'en ASCII.

Notez que notre RFC ne décrit pas un tel profil, il propose sa création, et donne quelques idées. Il y a donc encore du travail.

Pour rendre le problème plus amusant, les noms utilisés par DNS-SD sont composés de trois parties, avec des règles différentes (section 4 de notre RFC, et RFC 6763, section 4.1) :

— « Instance » qui est la partie présentée à l'utilisateur (« Imprimante de Céline »), et où donc tous les caractères sont acceptés,
— « Service », qui indique le protocole de transport, et qui se limite à un court identificateur technique en ASCII, jamais montré à l'utilisateur et qui, commençant par un tiret bas, n'est pas un nom de machine légal `<https://www.bortzmeyer.org/host-vs-domain.html>` et est donc exclu de tout traitement IDN,
— « Domaine », qui est le domaine où se trouve l'instance. Il est conçu pour être tôt ou tard envoyé au DNS, et il suit donc les règles IDN.

L'instance risque d'être « interceptée » par le traitement IDN et, à tort, transformée en Punycode. Limiter le nom de l'instance aux caractères acceptés par IDN serait horriblement restrictif. La seule solution est probablement de mettre en place du code spécifique qui reconnait les noms utilisés par DNS-SD, pour ne jamais les passer au DNS. Quant à la partie « Domaine », l'idée de la traiter comme un nom de domaine normal, et donc sujet aux règles des RFC 5890 et suivants, est restrictive :

— Cela revient à abandonner la stratégie de « on essaie UTF-8, puis Punycode » du RFC 6763, section 4.1.3 (mais cette stratégie a l'inconvénient de multiplier les requêtes, donc le trafic et les délais),
— Elle ne tient pas compte du fait que certains administrateurs de zones DNS peuvent avoir envie de mettre de l'UTF-8 directement dans la zone (ce qui est techniquement possible, mais va nécessiter de synchroniser dans la zone la représentation UTF-8 et la représentation Unicode du même nom, ou qu'il utilise les DNAME du RFC 6672).