RFC 7542 : The Network Access Identifier

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 2 mai 2015

Date de publication du RFC : Mai 2015

https://www.bortzmeyer.org/7542.html

Ce nouveau RFC normalise le concept d'**identificateur pour l'accès au réseau** (NAI pour "*Network Access Identifier*"). Cet identificateur est traditionnellement utilisé lors d'un accès authentifié au réseau, pour indiquer la personne ou l'entité qui veut se connecter (avec un mot de passe, ou autre information de créance, pour s'authentifier). Le NAI a une forme qui ressemble aux adresses de courrier ou XMPP. Par exemple, un NAI peut être `marie@example.com` ou `jean_dupont@labs.potamochère.fr`. Contrairement au traditionnel "*login*" ou nom d'utilisateur, le NAI inclut l'indication d'un domaine (ou plutôt royaume) d'origine, permettant au point d'accès d'accepter des utilisateurs issus d'un autre domaine. Le NAI permet donc des accès **fédérés**. Ce concept, en dépit de ce nom, est désormais utilisé pour bien d'autres choses que l'accès au réseau. Notre RFC remplace la norme précédente, le RFC 4282 [1], notamment en développant bien plus l'internationalisation (un NAI n'est pas forcément en ASCII). Bienvenue dans le monde merveilleux (et très bordélique) du AAA.

Initialement, la principale motivation pour le NAI ("*Network Access Identifier*") était le "*roaming*" : un utilisateur abonné au FAI X se déplace dans une zone où X ne fournit pas d'accès mais a un accord avec le FAI Y qui, lui, est présent. Pour éviter de recopier la base d'utilisateurs de X dans celle de Y, l'utilisateur présente un NAI qui indique son rattachement à X (son "*home domain*"). Le point d'accès de Y sait alors qu'il doit demander l'authentification à X. Une fois que c'est fait, on peut donner l'accès via Y (le "*visited domain*"). C'est par exemple ainsi que fonctionne la fédération Eduroam. (L'auteur du RFC est d'ailleurs l'auteur et le mainteneur de FreeRADIUS, un des logiciels les plus utilisés pour cette tâche. Le protocole RADIUS est normalisé dans le RFC 2865.)

Le NAI permet donc à des FAI purement régionaux d'accueillir des clients d'autres régions du pays, à des FAI nationaux d'accueillir des clients d'autres pays, à des "*hotspots WiFi*" de servir plusieurs FAI avec la même infrastructure, etc. Une description de l'usage des NAI dans des situations de "*roaming*" est donné dans le RFC 2194.

1. Pour voir le RFC de numéro NNN, `https://www.ietf.org/rfc/rfcNNN.txt`, par exemple `https://www.ietf.org/rfc/rfc4282.txt`

On l'a vu, le NAI peut servir à bien d'autres choses que l'accès au réseau. Sa définition est donc indépendante du protocole qui va l'utiliser (le NAI n'est pas spécifique à RADIUS.) Par contre, son encodage effectif dépendra du protocole (par exemple, il faudra assurer l'échappement de certains caractères, le NAI `sophie@internautique.fr` deviendra `sophie@internautique%2Efr` dans un URL). À noter que le NAI, quoique répandu, n'est pas **le** format unique d'identificateur sur le réseau. Certains protocoles existants ont un autre format, et certains permettent le NAI et d'autres formats. Pour les protocoles futurs, notre RFC recommande qu'ils adoptent le NAI, pour uniformiser les identificateurs. C'est déjà le cas de 3GPP dont la norme `<ftp://ftp.3gpp.org/Specs/archive/23_series/23.003/>` « *"TS 23.003 Numbering, addressing, and Identification (Release 12)*" » précise que l'identificateur est un NAI comme `234150999999999@ims.mnc015.mcc234.3gppnetwork.org`.

Attention, le but est d'uniformiser le format, pas les identificateurs. Le NAI n'implique pas que chaque utilisateur ait un identificateur et un seul ! Cela poserait, entre autres, de sérieux problèmes liés à la vie privée. Le RFC recommande d'ailleurs de permettre des identificateurs anonymes (pseudonymes, plutôt) dès qu'ils sont visibles publiquement.

Un peu de terminologie nécessaire pour comprendre le NAI figure en section 1.1. Notez le NAS ("*Network Access Server*") qui est la première machine à laquelle les utilisateurs se connectent pour avoir un accès à l'Internet. Pour les technologies PPTP et L2TP, ce sera le concentrateur d'accès. En WiFi, ce sera le point d'accès ("*hotspot*"). Lorsque l'utilisateur envoie son NAI, c'est le NAS qui extrait le domaine de l'utilisateur et relaie (par exemple en RADIUS) la demande d'authentification à ce domaine.

Une bonne partie des changements depuis le RFC 4282 avait été motivé par l'expérience d'Eduroam. Par exemple, la section 2.1 du RFC 4282 demandait que le nom de domaine soit uniquement composé de lettres ASCII, de chiffres et de tirets. Pour les IDN, l'idée était d'utiliser la forme Punycode. Celle-ci est peu pratique, et souvent inutile puisque plusieurs protocoles utilisant les NAI (comme RADIUS ou comme l'EAP du RFC 3748) ne sont pas limités à l'ASCII et recommandent UTF-8 comme encodage des chaînes de caractères (cf. section 3.2). Le RFC 4282 avait d'autres problèmes d'internationalisation comme d'exiger une normalisation des chaînes (qui peut rentrer en conflit avec des exigences locales) ou comme d'exiger des opérations qui soient dépendantes de la langue (que le NAS et autres équipements intermédiaires ne connaissent pas forcément, et ne savent pas toujours gérer). D'autre part, le RFC 4282 interdisait l'utilisation de points de code Unicode non affectés. Cela empêchait le déploiement de toute nouvelle écriture puisque, au début, tous les équipements réseau auraient considéré ces nouveaux caractères comme non affectés ! En pratique, d'ailleurs, aucun équipement réseau n'a mis en œuvre les recommandations d'internationalisation du RFC 4282. Le "*roaming*" international se développant, il était temps de changer ces recommandations irréalistes.

La section 2 de notre RFC présente la définition formelle du NAI. Il est en UTF-8, normalisé NFC. Il est divisé en deux parties, le nom d'utilisateur, et le royaume ("*realm*"). Ces deux parties sont séparées par un @. Le nom d'utilisateur peut être composée de lettres (Unicode, pas uniquement ASCII), de chiffres et de quelques symboles. `fred=?#$&*+-/{}smith` est donc un nom d'utilisateur valable. Le royaume ressemble à un nom de domaine, avec ses composants séparés par des points mais n'en est pas forcément un. Un exemple de NAI est donc `eng%geneviève@example.net` où `eng%geneviève` est le nom d'utilisateur et le domaine (royaume) est `example.net`. (La grammaire formelle est en section 2.2.) Un nom de royaume ne doit pas être réduit à un seul composant donc `maire@paris` n'est pas un NAI même si le TLD `.paris` existe. En effet, certains équipements considèrent qu'un royaume d'un seul composant est un sous-royaume du royaume local (donc, chez le FAI `example.net`, le NAI `maire@paris` serait interprété `maire@paris.example.net`).

Le royaume peut être en Unicode, bien des protocoles AAA autorisent à la transporter nativement et il est donc désormais déconseillé d'encoder en Punycode (RFC 3492). Ainsi, le NAI `faïza@café.example`

s'écrit bien ainsi, et pas `faïza@xn--caf-dma.example`. (Il semble bien que c'est ce que faisaient déjà tous les équipements qui géraient des noms de royaumes en Unicode, malgré ce que disait le RFC 4282.)

Il n'y a pas de limite de taille aux NAI, juste la recommandation de pouvoir gérer au moins 72 octets et, de préférence, 253. (Attention, en UTF-8, un caractère ne fait pas forcément un octet.) Les NAI étant utilisés dans des protocoles très différents, on a parfois des surprises. Ainsi, l'attribut `User-Name` de RADIUS exige d'accepter jusqu'à 63 octets mais ne dit rien au delà (RFC 2865, section 5.1). En revanche, le protocole concurrent Diameter (RFC 6733) impose à ses mises en œuvre des noms d'utilisateur jusqu'à 16 777 207 octets. Si on est sûr de toujours passer par Diameter (ce qui est très peu probable), on peut utiliser des NAI très longs.

Le nom d'utilisateur est opaque aux autres royaumes (comme pour le courrier électronique même si beaucoup d'ignorants <`https://www.bortzmeyer.org/arreter-d-interdire-des-adresses-legales.html`> violent cette règle). Par exemple, dans le `eng%geneviève@example.net` donné plus haut, on peut soupçonner que le nom d'utilisateur indique un routage interne (vers le service d'ingéniérie puis vers l'utilisatrice Geneviève) mais on ne peut pas en être sûr, chaque royaume a ses propres règles. On doit donc traiter le nom d'utilisateur comme un tout, sauf si on est le "*home domain*". Parfois, des protocoles transmettent juste le nom de royaume, faisant passer le nom d'utilisateur dans un canal plus sûr, afin de préserver la vie privée (cf. le TTLS du RFC 5281). L'habitude dans ce cas est de remplacer le nom d'utilisateur par `anonymous` mais c'est désormais déconseillé, il vaut mieux un nom vide (`@internautique.fr` est donc un NAI valide). Avec les autres protocoles, notre RFC recommande d'utiliser des noms d'utilisateurs éphémères, pour la vie privée, mais c'est très rarement le cas aujourd'hui.

On a vu qu'on pouvait utiliser des caractères Unicode aussi bien pour le nom de royaume que pour celui de l'utilisateur. Pour les noms d'utilisateur, les règles se basent sur celles du RFC 6532. Pour le nom de royaume, on est restreint aux caractères utilisables dans un nom de domaine (RFC 5891). L'éventuelle traduction en UTF-8 (au cas où l'utilisateur ait rentré des caractères dans un autre encodage) et la normalisation en NFC doivent être faites au tout début, lorsqu'on est encore proche de l'utilisateur et qu'on sait ce qu'il veut (on connait sa langue, par exemple, ce qui n'est pas le cas des équipements réseau intermédiaires ; ces équipements intermédiaires ont tout intérêt à ne **pas** tripoter les NAI). Ceci dit, cette situation idéale (les terminaux normalisent, les intermédiaires ne changent rien) n'est pas respectée aujourd'hui. Il existe des terminaux qui ne normalisent pas, injectant ainsi des chaînes non-UTF8 dans le système d'authentification. Le RFC note donc que le principe « les terminaux normalisent, les intermédiaires ne changent rien » doit parfois être violé. « "*The suggestion in the above sentence contradicts the suggestion in the previous section. This is the reality of imperfect protocols.*" » Ce point a été l'un des plus disputés lors de l'écriture de ce RFC.

Une fois le NAI défini, notre RFC s'attaque au routage des requêtes (section 3). Typiquement, le système d'authentification et autorisation (AAA) extrait le royaume du NAI et s'en sert comme clé pour une table où sont stockés les royaumes qu'on connait et avec qui on a une relation d'acceptation de leurs utilisateurs. Si le royaume n'est pas trouvé dans la table, on refuse l'utilisateur. S'il est trouvé, le contenu de la table nous dira le serveur à interroger pour ce royaume (ainsi que des informations comme le secret partagé RADIUS, le port, etc). Attention, la sémantique des noms de royaume n'est pas forcément connue et, par exemple, les équipements réseau ne savent pas si on peut les consulter de manière insensible à la casse. Parfois, ça marche (rappelez-vous le paragraphe précédent : le monde de l'AAA est imparfait...) Si on ne trouve pas un nom de royaume dans la table, on peut router sur une partie du royaume, par exemple utiliser `example.net` si le royaume `france.example.net` n'a pas été trouvé. Attention, ce n'est valable que si le nom raccourci reste un nom valide (`net` ne le serait pas, vu la prohibition des noms d'un seul composant).

Les NAI ressemblent aux adresses de courrier électronique et certaines normes sont communes aux deux (comme le RFC 6532) mais attention, les règles ne sont pas exactement les mêmes. Toute adresse

de courrier n'est donc pas forcément un NAI valide. En pratique, les deux se ressemblent suffisamment pour que beaucoup de FAI utilisent l'adresse de courrier du client comme son NAI.

On a vu que l'AAA était un monde très riche, ancien, et qui contient donc plein de choses surprenantes et de traditions historiques. La section 3.3.1 contient quelques exemples rigolos, avec les recommandations actuelles. Par exemple, les utilisateurs de RADIUS ont longtemps utilisé du routage explicite, où une partie du NAI contenait d'autres instructions de routage (`chezmoi.example!utilisateur@fai.exam`
Cette méthode (citée par le RFC 4282, section 2.7) s'est avérée très fragile, cassant dès qu'on change le réseau. RADIUS n'ayant pas de protocole de routage (qui diffuserait automatiquement les informations de routage), il fallait informer beaucoup de systèmes en cas de changement. (Diameter - RFC 5729 - ou 3G sont des cas différents car ils utilisent un protocole de routage.)

Le NAI étant en général utilisé dans un contexte de sécurité (point d'entrée pour une authentification), notre RFC consacre une section, la 4, à ces problèmes. Par exemple, si le protocole transporte le nom d'utilisateur en clair (c'est le cas de RADIUS), un écoutant sur le trajet peut apprendre des noms d'utilisateurs existants. Pour empêcher cela, il faut protéger (RADIUS avec IPsec, RFC 3579, Diameter avec TLS, RFC 6733). Si on utilise plusieurs protocoles qui tous se servent de NAI, et que l'utilisateur n'a qu'un seul NAI, un observateur pourra relier entre elles ces différentes utilisations. D'où l'intérêt de ne pas transporter les NAI en clair. Dans le futur, il serait encore mieux d'avoir des NAI éphémères, changés de temps en temps.

Le monde de l'accès réseau étant compliqué, il y a également des risques liés au fait qu'un identificateur peut être interprété d'une façon par un protocole et d'une autre façon par un autre protocole.

Dernier problème des NAI à régler, celui de l'avitaillement (création, maintenance, suppression) des identificateurs. Les noms de royaumes ressemblent à des noms de domaine pour éviter d'avoir à créer une nouvelle infrastructure d'avitaillement. Ainsi, si on est déjà titulaire du domaine `lesrépublicains.fr`, on n'a pas besoin d'autre chose pour créer des NAI comme `nicolas@lesrépublicains.fr`, tout en étant sûr de leur unicité, sur laquelle repose le routage. Bien qu'on puisse toujours, techniquement parlant, prendre un nom, sans l'enregistrer comme nom de domaine, pour faire des NAI, cette pratique est interdite. Par contre, l'usage du DNS n'est pas obligatoire et ces noms n'ont donc pas besoin d'être publiés. Si Diameter permet d'utiliser le DNS pour localiser un serveur d'authentification, les autres protocoles n'ont pas forcément cette possibilité : RADIUS repose sur des configurations statiques.

L'annexe A résume les (importants) changements depuis le RFC 4282, notamment :
— UTF-8 autorisé dans le nom de royaume,
— Abandon de Punycode.
Ces deux changements sont déjà largement déployés dans les équipements qui gèrent des NAI en Unicode.