RFC 7620 : Scenarios with Host Identification Complications

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 4 septembre 2015

Date de publication du RFC : Août 2015

https://www.bortzmeyer.org/7620.html

Un problème intéressant de l'Internet d'aujourd'hui est l'**identification** d'une machine ("*Host Identification*", ou "*Host-ID*", en anglais). Il s'agit de détecter si une machine A avec qui on communique est la même que la machine B avec qui on échange également. (Rappel : identification n'est pas authentification.) Évidemment, l'adresse IP ne convient plus depuis longtemps : les techniques de partage d'adresses IP, comme le CGN, sont trop répandues. Comment faire, alors ? Ce nouveau RFC ne propose pas de solutions mais il fait le tour d'horizon de ce problème complexe, plus complexe qu'on pouvait croire en lisant les solutions qui avaient été proposées par le RFC 6967 [1].

En effet, TCP/IP n'a pas de notion d'identité d'une machine. Au début, les adresses IP, stables et uniques, pouvaient remplir tant bien que mal ce rôle. Mais c'est fini depuis longtemps, surtout en IPv4. (IPv6 n'est pas forcément la solution : je connais des sites qui font du NAT IPv6 en sortie. Voir la section 3 de notre RFC.) Le RFC 6269 décrit en détail ce **partage d'adresses IP** et ses conséquences (en général néfastes pour l'identification). Notez que NAT et CGN ne sont pas les seules techniques qui amènent à un tel partage (par exemple, un relais applicatif a les mêmes conséquences : plusieurs machines derrière une seule adresse IP, cf. section 5 du RFC.). Et, en plus du partage d'adresses, les tunnels compliquent également les choses, puisqu'une machine aura une adresse très éloignée de son point d'attachement physique.

Outre les points mentionnés dans le RFC 6269, l'absence d'identification facile rend difficile :
— D'appliquer une politique spécifique à chaque machine (comme une limitation de trafic),
— De retrouver une machine interne quand on ne connait que l'adresse vue à l'extérieur,
— D'appeler un serveur de localisation, qui donne la position physique d'une machine, ce qui serait bien utile pour les appels d'urgence en VoIP (section 9 de notre RFC, et RFC 6443).

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc6967.txt

Évidemment, le partage d'adresses a aussi des avantages, comme une certaine dissimulation, utile pour la protection de la vie privée (cf. section 12 du RFC). Les solutions d'identification d'une machine peuvent potentiellement annuler certains efforts de protection de la vie privée <https://www.eff.org/issues/open-wireless>. Le RFC qui proposait des solutions, le RFC 6967 avait un principe : ne pas transmettre davantage d'informations que celles qui étaient dans le paquet IP original émis par la machine. Notre nouveau RFC se demande (section 2) si ce principe n'est pas trop strict.

La suite du RFC est constituée d'études de cas pour différents environnements. Par exemple, la section 3 étudie le classique CGN (RFC 6888). Certains cas seront résolus par IPv6 (par exemple le NAT IPv4 courant aujourd'hui) mais d'autres vont rester comme la traduction d'adresses IPv6 mentionnée plus haut (RFC 6296). La section 5 regarde les relais applicatifs : comme le CGN, ils « cachent » plusieurs machines derrière une même adresse IP. Comme pour le CGN, cela empêche certains usages comme une identification de l'abonné individuel, ou comme l'attribution d'une attaque, par exemple un commentaire menaçant sur un forum. On peut noter que des solutions spécifiques à tel ou tel protocole applicatif ont été développées et que, si le relais les utilise, le problème disparait. C'est le cas de l'en-tête `Received:` de SMTP (RFC 5321, sections 3.7.2 et 4.4), largement déployé. Dans le monde HTTP, il y a l'en-tête `Forwarded:`, du RFC 7239, moins fréquemment rencontré. Ces solutions sont souvent discutées dans le contexte de la protection de la vie privée car leur caractère indiscret est assez évident.

Autre étude de cas, en section 7, les "*overlays*". Ils ont beaucoup d'usages comme par exemple dans certaines architectures pair à pair (RFC 5694). Les manipulations qu'ils utilisent peuvent également aboutir à masquer les adresses IP d'origine. Le RFC contient également des études de cas sur d'autres architectures, le tout étant synthétisé dans la section 11 sous forme d'un joli tableau qui indique notamment, pour chaque architecture, si les problèmes qu'elle pose à l'identification d'une machine disparaitront ou pas avec IPv6.

Terminons en revenant sur la question de la protection de la vie privée car, évidemment, les mécanismes d'identification de la machine visent à envoyer des informations, là où la vie privée exigerait qu'on en envoie le moins possible. La section 12 du RFC expose le problème et renvoie à la section 3 du RFC 6967. Elle note aussi qu'il existe des « solutions » non-standard à la question de l'identification et que ces solutions sont souvent discutables en termes de vie privée (par exemple les en-têtes HTTP non-standards `HTTP_MSISDN`, `HTTP_X_MSISDN`, `HTTP_X_UP_CALLING_LINE_ID`, `HTTP_X_NOKIA_MSISDN`..) L'IESG a en outre ajouté une note au RFC, rappelant que, si on considère en effet les « problèmes » cités par ce RFC 7620 comme des problèmes à résoudre, les solutions ont de fortes chances de piétiner d'autres RFC et d'être contradictoires avec les principes de confidentialité des RFC 6280 et RFC 7258. Bref, identifier les machines ou protéger la vie privée des utilisateurs, il faudra sans doute choisir.