RFC 7721 : Privacy Considerations for IPv6 Address Generation Mechanisms

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 11 mars 2016

Date de publication du RFC : Mars 2016

https://www.bortzmeyer.org/7721.html

Il y a eu pas mal de trollage depuis les débuts d'IPv6, au sujet de son mécanisme de génération d'adresses, et des risques qu'il pouvait poser pour la vie privée. Ces débats étaient souvent motivés par une inquiétude de la nouveauté, davantage que par une réelle analyse des menaces. Plus d'un amateur a écrit sur les forums qu'IPv6 était un complot contre la vie privée, alors même que son navigateur Web faisait fuiter bien davantage d'informations personnelles. Ce nouveau RFC analyse l'état actuel des mécanismes de génération des adresses IPv6, leurs avantages et leurs inconvénients en terme de protection de la vie privée, et les compromis à faire lors du choix d'un mécanisme donné.

IPv6 permet des nouvelles façons d'allouer des adresses. Outre les méthodes qui existaient déjà en IPv4, comme l'allocation statique, et celle via DHCP, il permet une méthode dite « sans état ». Dans ce cas, le routeur annonce le préfixe IP du lien, et les machines ajoutent à ce préfixe un suffixe qui les identifie de manière unique sur le lien. Ainsi, on peut obtenir une adresse globale sans qu'un serveur ait besoin de maintenir un état des allocations faites.

Au début d'IPv6, des formulations malheurseuses dans les RFC ont pu faire croire que ce suffixe, l'IID ("*Interface IDentifier*") était forcément dérivé de l'adresse MAC de la machine, mais cela n'a jamais été obligatoire (le RFC 7136 [1] a clarifié ce point). D'autre part, on peut noter que préfixe et suffixe ont en général la même longueur, qui est de 64 bits, pour les raisons expliquées dans le RFC 7421.

Aujourd'hui, la liste complète des mécanismes possibles d'allocation d'adresses IPv6 est :

— Manuelle, en fixant l'IID en partant de 1 (méthode dite "*low byte*"), ou bien selon une adresse IPv4 de la machine, ou encore selon un port (serveurs DNS dont l'adresse IPv6 se termine par :53...), ou enfin en utilisant un mot rigolo, en profitant de l'hexspeak.

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc7136.txt

— Via la SLAAC, "*Stateless Address Auto-Configuration*", en fabriquant l'IID à partir de l'adresse MAC (RFC 2464), d'une clé cryptographique (RFC 3972), d'un nombre temporaire et donc changeant souvent (RFC 8981), ou d'une valeur stable dérivée de certaines caractéristiques du réseau d'accueil (RFC 7217). Et ce n'était que les méthodes normalisées, mais il y en a d'autres, par exemple chez Microsoft <http://www.microsoft.com/resources/documentation/windows/xp/all/proddocs/en-us/sag_ip_v6_imp_addr7.mspx?mfr=true>.
— Via DHCP (RFC 8415). Notez que le serveur DHCP peut avoir de nombreux moyens de choisir une adresse pour chaque client : cela peut être statique ou dynamique.
— Enfin, il y a les méthodes d'allocation des techniques de coexistence IPv4-IPv6. L'adresse v6 peut être dérivée de l'adresse v4 (RFC 6052), dérivée de l'adresse et du port (RFC 4380, dérivée de l'adresse et d'un identifiant d'un ensemble de ports (RFC 7596 et suivants)...

Parmi ces nombreuses méthodes, la méthode « SLAAC et IID dérivé de l'adresse MAC » est celle qui pose le plus de problèmes, côté vie privée. Pourquoi ? Parce que l'adresse MAC d'une machine est en général stable (sauf si on utilise un logiciel comme macchanger <http://www.gnu.org/software/macchanger>) même lorsque la machine se déplace. Si on se connecte à un pair sur l'Internet, depuis un certain réseau, le pair verra l'adresse IP source et, si on change de réseau, l'adresse MAC, donc l'IID, dans les 64 derniers bits de l'adresse IP, seront les mêmes et le pair verra donc qu'il s'agit de la même machine.

La section 3 de notre RFC décrit plus précisement les vulnérabilités de cette méthode. La première vulnérabilité est le risque de **corrélation**. Un pair avec qui on communique, ou bien un observateur du trafic réseau, peuvent se rendre compte que deux machines d'adresses IPv6 différentes sont en fait la même machine, en raison de l'égalité des deux IID. Comme on ne change pas de carte Ethernet tous les mois, la corrélation, dans le cas d'un IID fondé sur l'adresse MAC peut durer des années (cf. RFC 8981 pour l'analyse).

Bien sûr, on peut faire de la corrélation avec autre chose que les adresses MAC, comme l'illustre très bien le Panopticlick <https://panopticlick.eff.org/> (voir aussi la section 5.2.1 du RFC 6973) et une bonne partie des critiques contre ce risque de corrélation avec les IID liés aux adresses MAC étaient juste du FUD anti-IPv6 (des gens qui critiquent le manque de vie privée avec IPv6 mais qui ne filtrent pas les cookies...) Mais cela n'interdit pas d'essayer de résoudre le problème.

Un autre risque est celui de la localisation. Si une machine garde le même IID en changeant de réseau, on peut la suivre à la trace. « Tiens, c'est encore elle, et, cette fois, elle est en Allemagne. Tiens, elle est maintenant à l'université de Berlin. » Cette attaque n'aurait pas marché en IPv4 où l'adresse attribuée est complètement différente à chaque fois.

Les adresses IPv6 dérivées de l'adresse MAC présentent un autre risque : on peut balayer le réseau à la recherche de machines à attaquer. Normalement, la taille de l'espace d'adresssage IPv6 rend le balayage complet irréaliste. Mais, si on sait que le réseau visé n'utilise, par exemple, que des Apple, on peut réduire le nombre de paquets à envoyer en disant au logiciel de balayage de n'utiliser que des adresses IPv6 dérivées d'adresses MAC ayant le préfixe Apple. Et cette possibilité est mise en œuvre dans les outils réels, par exemple avec l'option `--tgt-ieee-oui` de scan6. (Voir le RFC 7707 pour plus de détail sur le balayage en IPv6.)

Enfin, le fait que les adresses MAC aient un préfixe qui indique le fabricant (sauf si on a pensé à utiliser macchanger <https://github.com/alobbs/macchanger> ou un logiciel équivalent) fournit une information qui peut être utile à un attaquant, en lui dévoilant le type de machines qu'on utilise.

Bien sûr, ces adresses IPv6 dérivées de l'adresse MAC sont minoritaires, et sont de moins en moins utilisées. Mais les alternatives ne sont pas toujours parfaites. La section 4 de notre RFC fait le point sur les solutions, et leurs propriétés. Les adresses temporaires du RFC 8981 sont la première solution à laquelle

on pense. Elles sont typiquement utilisées pour les connexions sortantes, une adresse stable, dérivée de l'adresse MAC, restant en place pour les connexions entrantes. À noter qu'au début, les problèmes de vie privée étant supposés moins importants que les considérations opérationnelles de gestion des réseaux, les RFC demandaient que les adresses temporaires ne soient pas utilisées par défaut. Plusieurs systèmes, comme Windows, ont décidé à juste titre d'ignorer cette demande, et c'est ce qui explique que les adresses dérivées d'une adresse MAC soient aujourd'hui minoritaires dans les journaux des serveurs Internet (voir les mesures à la fin de cet article). L'ancienne recommandation a été abandonné et le RFC 6724 dit clairement qu'il faut utiliser les adresses temporaires par défaut.

D'autre part, comme l'adresse stable fondée sur l'adresse MAC demeure, en plus de l'adresse temporaire, les inconvénients liés au balayage « optimisé » du réseau demeurent.

Il faut aussi noter qu'il n'y a pas que l'IID (la partie droite de l'adresse IP) qui est révélatrice. Le préfixe (la partie gauche) peut l'être également. Par exemple, si on suivait le RFC 3314, qui recommande un préfixe unique par téléphone en GPRS, ce préfixe, s'il était attribué de manière stable, serait caractéristique d'un téléphone donné.

Un tableau au début de la section 4 résume la vulnérabilité des différents choix d'allocation d'adresses IPv6 aux différents risques indiqués en section 3 (corrélation, localisation, balayage et indication du vendeur). Par exemple, une adresse attribuée manuellement permet la corrélation, tant que cette adresse dure (ces attributions manuelles concernent en général les serveurs, pour lesquels les problèmes de vie privée sont nettement moins graves) et une adresse dérivée de l'adresse MAC permet la corrélation tant qu'on ne change pas la carte Ethernet ou Wifi (ou son adresse MAC, ce que les gens oublient souvent, mais qui est possible avec la plupart des cartes).

En revanche, les adresses temporaires ne permettent la corrélation que pendant la durée de vie de ces adresses (une journée, par défaut), et les adresses stables et opaques du RFC 7217 permettent la corrélation tant que la machine ne change pas de réseau mais la bloquent dès qu'on bouge.

Le tableau est explicité dans les sections suivantes :

— Les adresses dérivées de l'adresse MAC sont sensibles aux quatre risques.
— Les adresses configurées manuellement sont sensibles aux deux premiers risques (corrélation et localisation).
— Des adresses générées cryptographiquement (par exemple en suivant le RFC 3972) sont en théorie vulnérables à la corrélation (puisqu'on garde la clé publique, et donc l'IID, très longtemps) mais cela peut être limité en suivant la section 7.3 du RFC 3972.
— Les adresses stables et opaques du RFC 7217 sont vulnérables à la corrélation tant qu'on reste sur le même réseau. Si une machine utilisant ces adresses change de réseau, il n'y a plus de corrélation possible mais, si elle revient sur le réseau d'origine, ses correspondants la reconnaitront à nouveau.
— Les adresses temporaires du RFC 8981 sont à l'abri des quatre vulnérabilités citées (ce qui est assez normal : elles sont conçues spécialement pour la vie privée). Il reste le cas du balayage : si l'adresse stable, utilisée pour les connexions entrantes, se fonde sur l'adresse MAC, certaines techniques de balayage peuvent encore s'utiliser. Il est donc recommandé d'utiliser le RFC 7217 ou équivalent pour générer l'adresse stable.

Enfin, la section 5 de notre RFC traite de divers problèmes liés à ce souci de vie privée. Par exemple, il rappelle que des adresses qui changent (telles les adresses temporaires du RFC 8981) peuvent être une plaie pour l'administration de réseaux et que certaines organisations vont donc les désactiver sur les postes de travail, et tant pis pour la vie privée.

Autre cas embêtant, il existe apparemment des suites de test IPv6 qui vérifient que l'adresse IPv6 est bien dérivée de l'adresse MAC. C'est évidemment une mauvaise idée et notre RFC demande que ces suites soient mises à jour.

À noter qu'il existe un outil pratique, `addr6`, qui fait partie du "*SI6 toolkit*" <`http://www.si6networks.com/tools/ipv6toolkit/`>. Utilisant divers techniques (et heuristiques : il se trompe parfois), il vous dit à quelle catégorie appartient une adresse :

```
[L'adresse de base d'une machine]
% addr6 -a 2a01:ff41:916b:3bb0:ba27:ebff:feaa:78b9
unicast=global=global=ieee-derived=b8-27-eb

[Une des adresses temporaires de la même machine (c'est un FreeBSD avec
ipv6_privacy="YES" dans rc.conf]
% addr6 -a 2a01:ff41:916b:3bb0:810e:9c6d:341:e7a
unicast=global=global=randomized=unspecified

[Un serveur public, d.nic.fr]
% addr6 -a 2001:678:c::1
unicast=global=global=low-byte=unspecified
```

Quelle est la situation aujourd'hui ? Quels types d'adresses sont utilisés par les machines IPv6 ? J'ai fait un petit test avec les visiteurs de mon blog (qui ne sont pas forcément représentatifs : vus les sujets traités sur ce blog, les visiteurs sont sans doute techniquement plus avancés que la moyenne) :

```
% zgrep -h -E  '^[0-9]+:' /var/log/apache2/access.log* | awk '{print $1}' | \
         sort | uniq | addr6 -i -s

** IPv6 General Address Analysis **

Total IPv6 addresses: 1997
Unicast:    1997 (100.00%)             Multicast:       0 (0.00%)
Unspec.:       0 (0.00%)

** IPv6 Unicast Addresses **

Loopback:           0 (0.00%)          IPv4-mapped:        0 (0.00%)
IPv4-compat.:       0 (0.00%)          Link-local:         0 (0.00%)
Site-local:         0 (0.00%)          Unique-local:       0 (0.00%)
6to4:               9 (0.45%)          Teredo:             0 (0.00%)
Global:          1988 (99.55%)

** IPv6 Interface IDs **

Total IIDs analyzed: 1997
IEEE-based:     230 (11.52%)           Low-byte:           323 (16.17%)
Embed-IPv4:       0 (0.00%)            Embed-IPv4 (64):    122 (6.11%)
Embed-port:       7 (0.35%)            Embed-port (r):       1 (0.05%)
ISATAP:           1 (0.05%)            Byte-pattern:        23 (1.15%)
Randomized:    1190 (59.59%)
```

On voit que la majorité utilise des adresses temporaires ("*Randomized*") ce qui est bon signe (et tord le cou à la légende comme quoi les adresses IPv6 permettraient de suivre un utilisateur à la trace). Le second type, en nombre d'adresses distinctes (je n'ai pas cherché à pondérer par le nombre de visites faites par chaque adresse), est celui des adresses manuelles avec un octet de faible valeur ("*Low-byte*"). Il peut s'agir de serveurs, par exemple des "*crawlers*". Les adresses IPv6 fondées sur l'adresse MAC sont le troisième type, avec plus de 11 %, et des efforts seront sans doute nécessaires pour sensibiliser ces utilisateurs.