RFC 8117 : Current Hostname Practice Considered Harmful

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 12 mars 2017

Date de publication du RFC : Mars 2017

https://www.bortzmeyer.org/8117.html

« Je suis l'iPhone de Jean-Luc! » Traditionnellement, les ordinateurs connectés à l'Internet ont un nom, et ce nom est souvent annoncé à l'extérieur par divers protocoles. Cette pratique très répandue, dont l'origine remonte à l'époque où on n'avait que quelques gros serveurs partagés, et fixes, est dangereuse pour la vie privée, dans un monde de mobilité et de machines individuelles. Comme le note ce nouveau RFC, « c'est comme si on se promenait dans la rue avec une étiquette bien visible portant son nom ». Ce RFC dresse l'état des lieux, fait la liste des protocoles problématiques, et suggère, lorsqu'on ne peut pas changer le protocole, d'utiliser des noms aléatoires, ne révélant rien sur la machine.

Pour illustrer le problème, voici un exemple du trafic WiFi pendant une réunion, en n'écoutant qu'un seul protocole, mDNS (RFC 6762 [1]). Et d'autres protocoles sont tout aussi bavards. Notez que cette écoute n'a nécessité aucun privilège particulier sur le réseau (cf. RFC 8386), ni aucune compétence. N'importe quel participant à la réunion, ou n'importe quelle personne située à proximité pouvait en faire autant avec tcpdump (j'ai changé les noms des personnes) :

```
% sudo tcpdump -n -vvv port 5353
tcpdump: listening on wlp2s0, link-type EN10MB (Ethernet), capture size 262144 bytes
15:03:16.909436 IP6 fe80::86a:ed2c:1bcc:6540.5353 > ff02::fb.5353: 0*- [0q] 2/0/3 0.4.5.6.C.C.B.1.C.2.D.E.A.6.8.
15:03:17.319992 IP 172.25.1.84.5353 > 224.0.0.251.5353: 0*- [0q] 2/0/3 C.4.1.6.F.8.D.E.0.3.6.3.4.1.0.1.0.0.0.0.0
15:03:20.699557 IP6 fe80::e2ac:cbff:fe95:da80.5353 > ff02::fb.5353: 0 [5q] [4n] [1au] PTR (QU)? _googlecast._tcp
```

1. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc6762.txt

On y voit que les noms des machines présentes sont annoncés à tous (`ff02::fb` et `224.0.0.251` sont des adresses "*multicast*"). Certains noms sont très révélateurs (nom, prénom et type de la machine), d'autres un peu moins (prénom et type), d'autres sont presques opaques (juste un type de machine, très général). Un indiscret qui regarde le trafic sur des réseaux publiquement accessibles peut ainsi se faire une bonne idée de quelles machines sont présentes, voire de qui est présent. Les deux exemples des noms `info-mac-66` et `John-Smiths-iPhone-7` illustrent les deux risques. Dans le premier cas, si le nom est stable, il permet de suivre à la trace une machine qui se déplacerait. Le second cas est encore pire puisqu'on a directement le nom du propriétaire.

Le fait que les ordinateurs aient des noms est une tradition très ancienne (voir la définition de "*host name*" dans le RFC 8499). Un nom court (sans point à l'intérieur) combiné avec un suffixe forme un FQDN ("*Fully Qualified Domain Name*", cf. RFC 1983). On utilise ces noms courts et ces FQDN à plein d'endroits. IP lui-même n'utilise pas de noms du tout mais plein de protocoles de la famille TCP/IP le font, comme mDNS montré plus haut.

Un nom court doit être unique dans un contexte donné mais n'est pas forcément unique mondialement. Le FQDN, lui, est unique au niveau mondial.

Je vous recommande l'excellent travail de M. Faath, F. Weisshaar et R. Winter, « "*How Broadcast Data Reveals Your Identity and Social Graph*" <`http://net.hs-augsburg.de/docs/paper_trac_2016.pdf`> » à l'atelier TRAC 2016 <`http://conferences.telecom-bretagne.eu/trac2016/`> (supports de leur exposé <`http://net.hs-augsburg.de/docs/paper_trac_2016_slides.pdf`>), montrant toutes les fuites d'information liées à cette utilisation des noms, et ce qu'un méchant peut en faire. (C'est ce groupe qui avait écouté le trafic WiFi lors d'une réunion IETF à Prague, déclenchant une grande discussion sur les attentes en matière de vie privée quand le trafic est diffusé.)

Pourquoi nomme-t-on les ordinateurs, au fait, à part la tradition ? Sur un réseau, bien des systèmes d'exploitation, à commencer par Unix et Windows tiennent pour acquis que les ordinateurs ont un nom, et ce nom peut être utilisé dans des tas de cas. Il existe plusieurs schémas de nommage (section 2 du RFC), du plus bucolique (noms de fleurs) au plus français (noms de vins) en passant par les schémas bien geeks comme les noms des personnages du Seigneur des Anneaux. Mais, parfois, c'est le système d'exploitation lui-même qui nomme l'ordinateur, en combinant le nom de l'utilisateur et les caractéristiques de l'ordinateur, comme on le voit avec les iPhones dans l'exemple tcpdump ci-dessus. (Sur les schémas de nommage, voir le RFC 1178, et, sur un ton plus léger, le RFC 2100. Il existe une excellente page Web pleine d'idées de noms <`http://seriss.com/people/erco/unixtools/hostnames.html`>. L'ISC fait des statistiques sur les noms vus <`https://ftp.isc.org/www/survey/reports/`> sur Internet. Entre 1995 <`https://ftp.isc.org/www/survey/reports/1995/01/firstnames.html`> et 2017 <`https://ftp.isc.org/www/survey/reports/2017/01/first.txt`>, vous pouvez constater la décroissance des noms sympas en faveur des noms utilitaires.)

Dans les environnements "*corporate*", on ne laisse pas l'utilisateur choisir et il y a un schéma officiel. Par exemple, sur le réseau interne de Microsoft, le nom est dérivé du nom de "*login*" de l'utilisateur et un des auteurs du RFC a donc une machine `huitema-test-2`.

Est-il nécessaire de donner des noms aux « objets », ces machines à laver ou brosses à dents connectés, qui sont des ordinateurs, mais ne sont en général pas perçus comme tels (ce qui a des graves conséquences en terme de sécurité) ? Comme ces engins n'offrent en général pas de services, ils ont moins besoin d'un nom facile à retenir, et, lorsque les protocoles réseaux employés forcent à utiliser un nom, c'est également un nom fabriqué à partir du nom du fabricant, du modèle de l'appareil et de son numéro de série (donc, un nom du genre `BrandX-edgeplus-4511-2539`). On voit même parfois la langue parlée par l'utilisateur utilisée dans ce nom, qui est donc très « parlant ».

Même un identificateur partiel peut être révélateur (section 3 du RFC). Si on ordinateur se nomme `dthaler-laptop`, on ne peut pas être sûr qu'il appartienne vraiment au co-auteur du RFC Dave Thaler. Il y a peut-être d'autres D. Thaler dans le monde. Mais si on observe cet ordinateur faire une connexion au réseau interne de Microsoft (pas besoin de casser le chiffrement, les métadonnées suffisent), on est alors raisonnablement sûr qu'on a identifié le propriétaire.

Beaucoup de gens croient à tort qu'un identificateur personnel doit forcément inclure le nom d'état civil de l'utilisateur. Mais ce n'est pas vrai : il suffit que l'identificateur soit stable, et puisse être relié, d'une façon ou d'une autre, au nom de l'utilisateur. Par exemple, si un ordinateur portable a le nom stable `a3dafaaf70950` (nom peu parlant) et que l'observateur ait pu voir une fois cette machine faire une connexion à un compte IMAP `jean_dupont`, on peut donc associer cet ordinateur à Jean Dupont, et le suivre ensuite à la trace.

Ce risque est encore plus important si l'attaquant maintient une base de données des identifications réussies (ce qui est automatisable), et des machines associées. Une ou deux fuites d'information faites il y a des mois, voire des années, et toutes les apparitions ultérieures de cette machine mèneront à une identification personnelle.

Donc, n'écoutez pas les gens qui vous parleront d'« anonymat » parce que les noms de machine ne sont pas parlants (comme le `a3dafaaf70950` plus haut). Si quelqu'un fait cela, cela prouve simplement qu'il ne comprend rien à la sécurité informatique. Un nom stable, pouvant être observé (et on a vu que bien des protocoles étaient très indiscrets), permet l'observation, et donc la surveillance.

Justement, quels sont les protocoles qui laissent ainsi fuiter des noms de machine, que l'observateur pourra noter et enregistrer (section 4 du RFC) ? Il y a d'abord DHCP, où le message de sollicitation initial (diffusé à tous…) contient le nom de la machine en clair. Le problème de vie privée dans DHCP est analysé plus en détail dans les RFC 7819 et RFC 7824. Les solutions pour limiter les dégâts sont dans le RFC 7844.

Le DNS est également une cause de fuite, par exemple parce qu'il permet d'obtenir le nom d'une machine à partir de son adresse IP, avec les requêtes PTR dans `in-addr.arpa` ou `ip6.arpa`, nom qui peut réveler des détails. C'est le cas avec tout protocole conçu justement pour distribuer des informations, comme celui du RFC 4620 (qui ne semble pas très déployé dans la nature).

Plus sérieux est le problème de mDNS (RFC 6762), illustré par le tcpdump montré plus haut. Les requêtes sont diffusées à tous sur le réseau local, et contiennent, directement ou indirectement, les noms des machines. Même chose avec le "*DNS Service Discovery*" du RFC 6763 et le LLMNR du RFC 4795 (beaucoup moins fréquent que mDNS).

Enfin, NetBIOS (quelqu'un l'utilise encore ?) est également une grande source d'indiscrétions.

Assez décrit le problème, comment le résoudre (section 5) ? Bien sûr, il faudra des protocoles moins bavards, qui ne clament pas le nom de la machine à tout le monde. Mais changer d'un coup des protocoles aussi répandus et aussi fermement installés que, par exemple, DHCP, ne va pas être facile. De même, demander aux utilisateurs de ne pas faire de requêtes DHCP lorsqu'ils visitent un réseau « non sûr » est difficile (déjà, comment l'utilisateur va-t-il correctement juger si le réseau est sûr ?), d'autant plus qu'ils risquent fort de ne pas avoir de connectivité du tout, dans ce cas. Certes, couper les protocoles non nécessaires est un bon principe de sécurité en général. Mais cet angle d'action semble quand même bien trop drastique. (Il faut aussi noter qu'il existe des protocoles privés, non-IETF, qui peuvent faire fuire des noms sans qu'on le sache. Le client Dropbox diffuse à la cantonade l'ID du client, et celui

des "*shares*" où il se connecte. Il est facile de faire un graphe des utilisateurs en mettant ensemble ceux qui se connectent au même "*share*".)

La suggestion de notre RFC est donc d'attaquer le problème d'une autre façon, en changeant le nom de la machine, pour lui substituer une valeur imprévisible (comme le fait le RFC 7844 pour les adresses MAC). Pour chaque nouveau réseau où est connectée la machine, on génère aléatoirement un nouveau nom, et c'est celui qu'on utilisera dans les requêtes DHCP ou mDNS. Ces protocoles fonctionneront toujours mais la surveillance des machines mobiles deviendra bien plus difficile. Bien sûr, pour empêcher toute corrélation, le changement de nom doit être coordonné avec les changements des autres identificateurs, comme l'adresse IP ou l'adresse MAC.

Windows a même un concept de « nom de machine par réseau », ce qui permet aux machines ayant deux connexions de présenter deux identités différentes (malheureusement, Unix n'a pas ce concept, le nom est forcément global).

Bien sûr, on n'a rien sans rien (section 6). Si on change les noms des machines, on rendra l'administration système plus difficile. Par exemple, l'investigation sur un incident de sécurité sera plus complexe. Mais la défense de la vie privée est à ce prix.

Pour l'instant, à ma connaissance, il n'y a pas encore de mise en œuvre de cette idée de noms imprévisibles et changeants. (Une proposition a été faite <https://labs.riseup.net/code/issues/7061> pour Tails. Notez qu'il existe d'autres possibilités comme d'avoir un nom unique partout <https://mailman.boum.org/pipermail/tails-dev/2013-January/002457.html>.)