Le bitsquatting menace-t-il les utilisateurs de l'Internet ?

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 23 mai 2013

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

Vous avez confiance dans les composants électroniques de votre ordinateur ? Vous croyez que, lorsque vous tapez ou sélectionnez http://www.impots.gouv.fr/, vous vous retrouverez bien là où vous pouvez déclarer vos revenus ? Vous avez tort. L'électronique, c'est fragile. Un bit 1 peut se changer en 0 subitement (ou le contraire) et, à partir de là, le http://kremlin.ru/ se change en http://kremlin.re/... C'est le "*bit flipping*" (le changement de la valeur d'un bit), qui permet le "*bitsquatting*" (l'enregistrement par un méchant d'un domaine proche d'un bit du domaine dont on veut détourner le trafic).

Mais pourquoi un 0 se transformerait-il en 1 subitement ? Il existe des tas de raisons physiques possibles, d'un rayon cosmique (la Terre en reçoit en permanence) à la radioactivité en passant par la simple agitation thermique. Moins il y a d'atomes pour faire un bit (avec les progrès de la miniaturisation) et plus le risque est important. Mais les ordinateurs n'ont pas de moyen de se défendre, de la redondance, des contrôles ? Si, cela existe, cela se nomme les mémoires ECC mais elles sont plus chères et n'équipent pas les engins de bas de gamme. Ceux-ci, bon marché mais connectés à l'Internet, sont de plus en plus nombreux.

Quels sont les changements possibles ? En regardant la table ASCII, on voit que e est représenté par 01100101 et u par 01110101. Un changement du quatrième bit suffit donc à passer de la Russie à l'île de la Réunion comme dans l'exemple ci-dessus. Mais le "*bit flipping*" ne change pas que des lettres en lettres. Par exemple, n est représenté par 01101110 et le point par 00101110. En changeant le deuxième bit, on change tout dans le nom de domaine. windowsupdate.com peut devenir wi.dowsupdate.com en modifiant un seul bit. Même chose entre o (01101111) et la barre oblique (00101111). Un https://ecampus.phoenix.edu tapé dans un navigateur peut devenir https://ecampus.ph/enix.edu, qui est dans un tout autre TLD.

Mais les protocoles de sécurité comme X.509 ou DNSSEC ne vont pas s'y opposer ? Si le "*bit flipping*" avait lieu dans le réseau, sans doute. Mais il a souvent lieu dans la machine originale. Auquel cas, ces protocoles ne peuvent pas aider, le problème étant dès le début. S'il a lieu dans le réseau, notons que les simples mécanismes de contrôle existants comme la somme de contrôle UDP suffisent à l'attraper.

Bon, le "*bit flipping*" est possible. Est-il fréquent ? Pose-t-il un vrai problème en pratique ? La première question est délicate car on ne connait pas le nombre de fois où un nom de domaine est manipulé et copié dans une machine. Même si le pourcentage de "*bit flipping*" est très faible, il faut le multiplier par le nombre de machines existantes et par le nombre de fois qu'elles tripotent des noms de domaine. En fait, personne ne sait vraiment.

Et le risque de sécurité ? C'est que quelqu'un de mal intentionné n'enregistre le nom avec un bit changé et n'intercepte alors du trafic légitime, comme dans l'exemple du Kremlin ci-dessus. L'expérience (voir la bibliographie) montre que ces noms "*bitsquattés*" attirent effectivement du trafic, même s'il n'est pas toujours facile d'être certain de son origine. À la dernière réunion OARC <https://WWW.DNS-OARC.NET/> à Dublin, Jaeson Schultz a présenté une entreprise de "*bitsquatting*" de grande envergure, afin d'étudier le phénomène, et ses mesures semblent indiquer que le "*bit flipping*" est plus répandu qu'on en le pensait.

Il a aussi étudié le passé et montré que wwwnfacebook.com, "*bitsquatting*" de www.facebook.com, avait été enregistré deux ans avant la publication du premier papier sur le "*bitsquatting*". Cela ne veut pas dire que celui qui a fait l'enregistrement connaissait le phénomène du "*bit flipping*", peut-être le "*domainer*" a-t-il essayé beaucoup de noms et constaté empiriquement que celui-ci recevait du trafic.

Mais il est difficile de faire la part de ce qui est du vrai "*bit flipping*". Il peut s'agir de fautes de frappe (Schultz note que des domaines très éloignés sur le clavier mais proches en bits reçoivent eux aussi du trafic, donc les fautes de frappe n'expliquent pas tout) ou d'un simple bruit de fond (enregistrez n'importe quel nom de domaine, vous aurez du trafic).

L'article de Schultz contient aussi des suggestions de techniques pour limiter le "*bitsquatting*" mais aucune ne me semble réaliste. J'en ai quand même déployé une dans le source de cet article : les noms de domaine dans les URL sont en majuscules (il y a moins de possibilités de "*bit flipping*" en majuscule).

Un peu de bibliographie :

- La première publication sur le bitsquatting <http://DINABURG.ORG/bitsquatting.html>, en 2011, montrant que le phénomène est réel.
- Une étude par VeriSign <http://WWW.VERISIGNINC.COM/assets/VRSN_Bitsquatting_TR_20120320.pdf> en 2012 sur la fréquence du "*bit flipping*" dans les requêtes reçues par les serveurs de noms de .com, qui montre que les sommes de contrôle UDP sont en général correctes (s'il a lieu, le "*bit flipping*" a donc lieu dans la machine de départ) et les transparents de l'auteur <https://WWW.DNS-OARC.NET/files/workshop-201110/observations-on-checksum-errors.pdf> lors d'un exposé à l'atelier OARC <https://WWW.DNS-OARC.NET/> de Vienne,
- Un précédent article en français <http://KORBEN.INFO/bitsquatting-comment-ca-marche.html> sur le "*bitsquatting*".
- Un bon article du Wikipédia anglophone sur les erreurs en mémoire,
- Une intéressante étude <http://www.slideshare.net/nicknikiforakis/bitsquatting-exploiti de 2013 sur le "*bitsquatting*" effectif : est-il réellement exploité ajourd'hui ?
- L'exposé <https://INDICO.DNS-OARC.NET/indico/getFile.py/access?contribId=5&resId=3&materialId=slides&confId=0> de Schultz à la dernière réunion OARC à Dublin. L'article correspondant n'est pas encore publié mais je l'ai lu et il est très intéressant et plein de détails.

Enfin, un script Python pour afficher les variantes "*bit-flipées*" d'un nom, (en ligne sur https://www.bortzmeyer.org/files/bitflip.py) :

```
% python bitflip.py  labanquepostale
mabanquepostale
nabanquepostale
habanquepostale
dabanquepostale
lcbanquepostale
lebanquepostale
libanquepostale
lqbanquepostale
lacanquepostale
...
```