Réseaux sociaux et historique de navigation, les ennemis de l'anonymat

L'anonymat sur Internet est, on le sait, une illusion. Le fonctionnement intrinsèque du réseau attache l'utilisateur (ou plus précisément son terminal) à une adresse IP exploitable (la plupart du temps) pour retrouver son identité. Quitte à mettre en action de longs et coûteux services juridiques dans les affaires les plus graves. Mais, sans même s'appuyer sur cet identifiant numérique unique, il reste possible d'identifier le profil d'un internaute en croisant simplement son historique de navigation avec les informations publiques de ses comptes de réseaux sociaux.

C'est en tout cas ce que prétendent pouvoir accomplir des chercheurs des universités de Stanford et Princeton. Dans leur travaux, ils déclarent avoir été en mesure d'identifier plus de 70% des près de 400 internautes qui se sont prêtés au jeu, dont 50% en s'appuyant sur leur seul compte Twitter. La méthode consiste à repérer les connexions entre les liens partagés sur le réseau social et la probabilité qu'un utilisateur privilégie des recommandations personnelles sur l'ensemble de sa navigation web.

Plus de 50% de succès

« Notre approche est basée sur une simple observation : chaque personne a un réseau social distinctif et, donc, l'ensemble des liens apparaissant dans son flux est unique, écrivent les scientifiques. *En supposant que les utilisateurs visitent les liens dans leur flux avec une probabilité plus élevée qu'un utilisateur aléatoire, les historiques de navigation contiennent des identités. Nous officialisons cette intuition en spécifiant un modèle comportemental de navigation sur le Web, puis en obtenant le maximum d'estimations vraisemblables du profil social d'un utilisateur. Nous évaluons cette stratégie d'historique de navigation simulée et montrons qu'avec 30 liens provenant de Twitter, nous pouvons déduire le profil Twitter [de l'utilisateur] plus de 50% du temps. »* Autrement dit : « Dis moi ce que tu twittes (ou suis), je te dirais qui tu es ».

Pour mener à bien leurs travaux, les chercheurs ont installé une extension dans Chrome pour recueillir l'historique de navigation des 374 volontaires de l'expérience. Ils ont ensuite utilisé le protocole propriétaire de raccourcissement d'URL de Twitter pour identifier les liens t.co. L'utilisateur était alors correctement identifié dans 81% des cas dans les 15 premières requêtes de désanonymisation. Un taux rapporté à 72% dès la première recherche d'identité. Appliqué à grande échelle, près des trois-quarts des internautes seraient ainsi identifiables.

Les risques du suivi en ligne

L'objet de ces travaux a notamment pour intérêt de pointer les risques que les utilisateurs qui souhaitent préserver une forme de vie privée en ligne encourent face aux traqueurs publicitaires et autres adversaires de l'anonymat. Certes, les sites web visités ne disposent pas de l'historique de navigation du visiteur (contrairement aux chercheurs pour leur expérience). Mais les annonceurs et autres courtiers de données disposent d'outils suffisamment évolués (cookies, etc.) pour retracer

cet historique. Bref, il ne serait pas surprenant que les techniques de ré-identification des internautes soit aujourd'hui exploitées par les sites intéressés.

Pour limiter les risques d'authentification, il reste donc à l'utilisateur la possibilité d'utiliser un profil social réellement anonyme (sans aucun identifiant personnel). Une tendance que les réseaux sociaux essaient de bannir tant pour des raisons de crédibilité que de sécurité. L'usage de réseau VPN et de proxy «anonymes» peut également compliquer la tâche des désanonymiseurs. Mais le suivi des internautes à coup de cookies et autres traces laissées en ligne permettent de contourner ces mesures de protection. *« La couche applicative de désanonymisation a longtemps été considérée comme le talon d'Achille de Tor et d'autres systèmes d'anonymat, et notre travail montre que c'est bien le cas »*, indiquent les scientifiques. Une raison supplémentaire pour vider régulièrement l'historique de sa navigation et supprimer tout aussi fréquemment les cookies enregistrés.

Lire également
Un gourou du chiffrement lance PrivaTegrity, une alternative à Tor
WiFi public : la justice européenne flingue l'anonymat sur l'autel du droit d'auteur
Les sites cachés Tor exposés grâce au serveur Apache

crédit photo © Realinemedia – Shutterstock