

Même refusés, les cookies « phénix » continuent de traquer les internautes

Description

La fin des cookies déposés dans les navigateurs web semble validée. Deux études montrent combien certains acteurs du web détournent allègrement la réglementation encadrant l'usage des cookies. Depuis 2020, Google tente de proposer une alternative aux cookies : Topics, la plus récente, est testée au niveau mondial.

Depuis les premiers comme NCSA Mosaic créé en 1993 et Netscape Navigator en 1995, les logiciels de navigation sur le web se sont considérablement complexifiés. Aujourd'hui, le navigateur Chrome développé par Google représente 66 % du marché mondial, suivi de Safari développé par Apple, avec 15 % du marché, puis de Firefox développé par Mozilla et de Edge + IE produit par Microsoft. Si ces navigateurs web ont grandement amélioré leur fluidité d'usage, bon nombre d'entreprises en ont aussi profité pour élaborer un modèle économique à partir du suivi des internautes. Une équipe de chercheurs issue d'une douzaine de laboratoires de recherche – parmi lesquels l'Inria (Institut national de recherche en informatique et en automatique), l'Insa Lyon (Institut national des sciences appliquées), l'Université d'Utrecht, ou encore l'équipe du Laboratoire d'innovation numérique de la Cnil (Commission nationale de l'informatique et des libertés), le LINC – a récemment publié la première étude permettant de « *détecter et mesurer la réapparition des cookies à l'aide de l'empreinte digitale du navigateur et de la machine* ».

Avec la mise en œuvre du règlement général sur la protection des données (RGPD) en mai 2018 ([voir La rem n°42-43, p.21](#)), quiconque offre l'accès à un site web doit laisser le choix à l'internaute d'accepter ou de refuser le dépôt de cookies sur son navigateur. Alors que les cookies dits « propriétaires » ou « internes » sont déposés exclusivement par le site web consulté par l'internaute, les cookies tiers sont, explique la Cnil, « *des cookies déposés sur des domaines différents de celui du site principal, généralement gérés par des tiers qui ont été interrogés par le site visité et non par l'internaute lui-même* ». Il s'agit notamment des cookies tiers attachés aux boutons de partage sur les réseaux sociaux, aux moteurs de recherche ou encore aux services de publicité en ligne, qui servent pareillement à suivre la navigation d'un internaute.

Deux études menées par cette équipe de chercheurs ont démontré comment le refus des cookies par un internaute est contourné par des entreprises peu scrupuleuses qui continuent de « tracer » les internautes. Ces cookies sont appelés « phénix », en référence à l'oiseau mythique dont le pouvoir est de « renaître de ses cendres ».

La première étude s'attache à analyser les sites web les plus visités au monde et la seconde concerne les sites de santé pour lesquels le traçage des visiteurs n'a pas uniquement un objectif publicitaire mais

s'applique également aux données sensibles. 1 150 des 30 000 premiers sites web classés par Alexa (entreprise qui fournissait des statistiques sur le trafic du web mondial et dont le propriétaire, Amazon, a fermé le site en mai 2022) déploient une stratégie « *pour suivre les utilisateurs sur plusieurs sites web même lorsque les cookies tiers sont refusés* ». Le procédé, mis en lumière par les chercheurs, consiste à créer une « *empreinte digitale du navigateur et de la machine* » en enregistrant des informations comme le système d'exploitation, le navigateur web et sa version, le fuseau horaire ou encore la langue utilisée. À partir de cette empreinte, associée au cookie initial, il sera possible de recréer celui-ci même s'il a été effacé. Après avoir identifié 400 000 cookies, les chercheurs ont modifié artificiellement l'empreinte des ordinateurs participant à l'étude afin de démontrer la réapparition des cookies, pourtant déjà effacés : 425 cookies provenant de 1 150 sites ont été qualifiés de « phénix ».

La deuxième étude porte exclusivement sur 385 sites web de médecins, d'hôpitaux, de prises de rendez-vous médicaux. Premier enseignement : 76 % des sites ne respectent pas le principe du consentement explicite imposé par le RGPD. En outre, 62 % des sites web identifient le visiteur avant même de proposer d'accepter ou de refuser les cookies, 40 % ne proposent pas de refuser le dépôt de cookies et 15 % continuent de le pister en dépit de son refus. La méthode utilisée par ces sites web repose sur les « pixels invisibles ». Lorsqu'un internaute visite un site web, une requête est envoyée vers un site tiers qui affiche alors sur la page un pixel blanc, impossible à visualiser, et dont l'objet sera de déposer un nouveau cookie ou de récupérer les informations d'un cookie déjà présent dans le navigateur.

Les abus et les innovations techniques détournant les cookies à des fins de traçabilité n'ont de secret pour personne. D'autres abus ont cours : une étude menée par des chercheurs de trois universités – néerlandaise, suisse et belge – a montré que lorsqu'un internaute remplit un formulaire en ligne, l'adresse e-mail de l'internaute est récupérée avant même qu'il ait cliqué sur le bouton « Envoyer ».

Google travaille depuis 2020 à une alternative aux cookies, avec la Privacy Sandbox, présentée comme étant plus respectueuse de la vie privée. L'entreprise californienne a d'abord envisagé une technique de suivi basée sur l'apprentissage fédéré des cohortes (*Federated Learning of Cohorts*, FLoC, voir *infra*). Ce système, censé concilier le respect de la vie privée et le ciblage publicitaire, aurait permis aux annonceurs de cibler non plus chaque individu mais des cohortes d'utilisateurs regroupant plusieurs centaines ou milliers d'internautes avec un profil et des intérêts semblables. Le lancement de FLoC, initialement prévu pour le début de l'année 2022, a d'abord été repoussé à mi-2023 face à la fronde de nombreux acteurs, puis finalement abandonné début 2022. L'Electronic Frontier Foundation (EFF), une ONG internationale de protection des libertés sur internet basée à San Francisco en Californie, les éditeurs des navigateurs Mozilla, Brave, Vivaldi ou encore DuckDuckGo avaient déjà manifesté leur refus d'intégrer cette technologie. Face à cette levée de bouclier, Google propose dorénavant une méthode appelée Topics qui, selon le géant du web, « *permettrait aux annonceurs de montrer des publicités pertinentes aux internautes en fonction de leur historique de navigation, sans pour autant révéler à des tiers, ni même à Google, les différents sites qu'ils ont visités* ». Avec Topics, selon l'historique de navigation conservé pendant trois semaines, le navigateur identifie les thèmes correspondant aux centres d'intérêt de l'internaute et transmet trois de ces thèmes aux annonceurs afin qu'ils ciblent leurs publicités. La méthode permettrait de traiter les données au sein même

de l'appareil, sans que la moindre information soit transmise ni à Google, ni à des tiers. Des tests de Topics sur le navigateur Chrome sont en cours au niveau mondial.

Rappelons que, selon GroupM (les agences média du groupe WPP), les trois géants du numérique, Google (Alphabet), Facebook (Meta) et Amazon, captent plus de 50 % des parts du marché mondial de la publicité (voir [La rem n°42-43, p.92](#) et [p.99](#)), estimé à 550 milliards de dollars en 2021. Tel David contre Goliath, des internautes, des artistes et des défenseurs des droits tentent de résister à ce pistage systématique en utilisant des techniques dites d'« obfuscation » (qui consiste à protéger sa vie privée sur internet) ou d'empoisonnement de données (*data poisoning*) dont l'objet est d'induire en erreur les algorithmes de ciblage en produisant et disséminant de fausses informations. Mais ces techniques chronophages, et dont la portée reste limitée, sont davantage des outils de sensibilisation que des contre-pouvoirs. Si la fin des cookies est dorénavant décidée, nul ne sait quelle forme prendra le suivi publicitaire en 2023, ni quelle technologie pourrait résoudre l'équation périlleuse conciliant le respect de la vie privée des internautes et les intérêts financiers du marché de la publicité en ligne, de plus en plus invasive.

Sources :

- « FLoC : comment la méthode de ciblage sans cookies de Google veut révolutionner la pub en ligne », Mathieu Chartier, [lesnumeriques.com](#), 14 avril 2021.
- « Publicité : Google, Facebook (Meta) et Amazon pèsent plus de 50 % du marché mondial », Nicolas Richaud, [lesechos.fr](#), 9 décembre 2021.
- « Vie privée : les internautes, même précautionneux, pistés à leur insu », David Larousserie, *Le Monde*, 13 avril 2022.
- « Contre la surveillance en ligne, des internautes « empoisonnent » leurs données personnelles », Aurélien Defer, *Le Monde*, 29 avril 2022.
- « Des milliers de sites Internet enregistrent les données des formulaires, avant même que le bouton « Envoyer » ait été utilisé », Maxime Pernet, [lemonde.fr](#), 12 mai 2022.
- « My Cookie is a phoenix : detection, measurement, and lawfulness of cookie respawning with browser fingerprinting », Imane Fouad, Cristiana Santos, Arnaud Legout, Nataliia Bielova, PETS 2022, 22nd Privacy Enhancing Technologies Symposium, Sydney, Australia, July 2022.

Categorie

1. Techniques

date créée

8 juillet 2022

Auteur

jacquesandrefines