

Google chante le requiem pour les cookies, mais le grand chœur du pistage résonnera encore

Google va cesser de nous pister avec des cookies tiers ! Une bonne nouvelle, oui mais... Regardons le projet d'un peu plus près avec un article de l'EFF.

La presse en ligne s'en est fait largement l'écho : par exemple siecledigital, generation-nt ou lemonde. Et de nombreux articles citent un éminent responsable du tout-puissant Google :

Chrome a annoncé son intention de supprimer la prise en charge des cookies tiers et que nous avons travaillé avec l'ensemble du secteur sur le Privacy Sandbox afin de mettre au point des innovations qui protègent l'anonymat tout en fournissant des résultats aux annonceurs et aux éditeurs. Malgré cela, nous continuons à recevoir des questions pour savoir si Google va rejoindre d'autres acteurs du secteur des technologies publicitaires qui prévoient de remplacer les cookies tiers par d'autres identifiants de niveau utilisateur. Aujourd'hui, nous précisons qu'une fois les cookies tiers supprimés, nous ne créerons pas d'identifiants alternatifs pour suivre les individus lors de leur navigation sur le Web et nous ne les utiliserons pas dans nos produits.

David Temkin, Director of Product Management, Ads Privacy and Trust (source)

« Pas d'identifiants alternatifs » voilà de quoi nous réjouir : serait-ce la fin d'une époque ?

Comme d'habitude avec Google, il faut se demander où est l'arnaque lucrative. Car il semble bien que le Béhémoth du numérique n'ait pas du tout renoncé à son modèle économique qui est la vente de publicité.

Dans cet article de l'Electronic Frontier Foundation, que vous a traduit l'équipe de Framalang, il va être question d'un projet déjà entamé de Google dont l'acronyme est FLoC, c'est-à-dire **Federated Learning of Cohorts**. Vous le trouverez ici traduit **AFC** pour « Apprentissage Fédéré de Cohorte » (voir l'article de Wikipédia Apprentissage fédéré).

Pour l'essentiel, ce dispositif donnerait au navigateur Chrome la possibilité de créer des groupes de milliers d'utilisateurs ayant des habitudes de navigation similaires et permettrait aux annonceurs de cibler ces « cohortes ».

Article original sur le blog de l'Electronic Frontier Foundation : Google's FLoC is a terrible idea.

Traduction Framalang : amostra, audionuma, Fabrice, goofy, jums, Mannik, mo, amostra, serici, Wisi_eu

Le FLoC de Google est une très mauvaise idée

par Bennett Cyphers



Les cookies tiers se meurent, mais Google essaie de créer leur remplaçant.

Personne ne devrait pleurer la disparition des cookies tels que nous les connaissons aujourd'hui. Pendant plus de deux décennies, les cookies tiers ont été la pierre angulaire d'une obscure et sordide industrie de surveillance publicitaire sur le Web, brassant plusieurs milliards de dollars ; l'abandon progressif des cookies de pistage et autres identifiants tiers persistants tarde à arriver. Néanmoins, si les bases de l'industrie publicitaire évoluent, ses acteurs les plus importants sont déterminés à retomber sur leurs pieds.

Google veut être en première ligne pour remplacer les cookies tiers par un ensemble de technologies permettant de diffuser des annonces ciblées sur Internet. Et certaines de ses propositions laissent penser que les critiques envers le capitalisme de surveillance n'ont pas été entendues. Cet article se concentrera sur l'une de ces propositions : l'Apprentissage Fédéré de Cohorte (AFC, ou FLoC en anglais), qui est peut-être la plus ambitieuse – et potentiellement la plus dangereuse de toutes.

L'AFC est conçu comme une nouvelle manière pour votre

navigateur d'établir votre profil, ce que les pisteurs tiers faisaient jusqu'à maintenant, c'est-à-dire en retravaillant votre historique de navigation récent pour le traduire en une catégorie comportementale qui sera ensuite partagée avec les sites web et les annonceurs. Cette technologie permettra d'éviter les risques sur la vie privée que posent les cookies tiers, mais elle en créera de nouveaux par la même occasion. Une solution qui peut également exacerber les pires attaques sur la vie privée posées par les publicités comportementales, comme une discrimination accrue et un ciblage prédateur.

La réponse de Google aux défenseurs de la vie privée a été de prétendre que le monde de demain avec l'AFC (et d'autres composants inclus dans le « bac à sable de la vie privée » sera meilleur que celui d'aujourd'hui, dans lequel les marchands de données et les géants de la tech pistent et profilent en toute impunité. Mais cette perspective attractive repose sur le présupposé fallacieux que nous devrions choisir entre « le pistage à l'ancienne » et le « nouveau pistage ». Au lieu de réinventer la roue à espionner la vie privée, ne pourrait-on pas imaginer un monde meilleur débarrassé des problèmes surabondants de la publicité ciblée ?

Nous sommes à la croisée des chemins. L'ère des cookies tiers, peut-être la plus grande erreur du Web, est derrière nous et deux futurs possibles nous attendent.

Dans l'un d'entre eux, c'est aux utilisateurs et utilisatrices que revient le choix des informations à partager avec chacun des sites avec lesquels il ou elle interagit. Plus besoin de s'inquiéter du fait que notre historique de navigation puisse être utilisé contre nous-mêmes, ou employé pour nous manipuler, lors de l'ouverture d'un nouvel onglet.

Dans l'autre, le comportement de chacune et chacun est répercuté de site en site, au moyen d'une étiquette, invisible à première vue mais riche de significations pour celles et ceux qui y ont accès. L'historique de navigation récent,

concentré en quelques bits, est « démocratisé » et partagé avec les dizaines d'interprètes anonymes qui sont partie prenante des pages web. Les utilisatrices et utilisateurs commencent chaque interaction avec une confession : voici ce que j'ai fait cette semaine, tenez-en compte.

Les utilisatrices et les personnes engagées dans la défense des droits numériques doivent rejeter l'AFC et les autres tentatives malvenues de réinventer le ciblage comportemental. Nous exhortons Google à abandonner cette pratique et à orienter ses efforts vers la construction d'un Web réellement favorable aux utilisateurs.

Qu'est-ce que l'AFC ?

En 2019, Google présentait son bac à sable de la vie privée qui correspond à sa vision du futur de la confidentialité sur le Web. Le point central de ce projet est un ensemble de protocoles, dépourvus de cookies, conçus pour couvrir la multitude de cas d'usage que les cookies tiers fournissent actuellement aux annonceurs. Google a soumis ses propositions au W3C, l'organisme qui forge les normes du Web, où elles ont été principalement examinées par le groupe de commerce publicitaire sur le Web, un organisme essentiellement composé de marchands de technologie publicitaire. Dans les mois qui ont suivi, Google et d'autres publicitaires ont proposé des dizaines de standards techniques portant des noms d'oiseaux : pigeon, tourterelle, moineau, cygne, francolin, pélican, perroquet... et ainsi de suite ; c'est très sérieux ! Chacune de ces propositions aviaires a pour objectif de remplacer différentes fonctionnalités de l'écosystème publicitaire qui sont pour l'instant assurées par les cookies.

L'AFC est conçu pour aider les annonceurs à améliorer le ciblage comportemental sans l'aide des cookies tiers. Un navigateur ayant ce système activé collecterait les informations sur les habitudes de navigation de son utilisatrice et les utiliserait pour les affecter à une

« cohorte » ou à un groupe. Les utilisateurs qui ont des habitudes de navigations similaires – reste à définir le mot « similaire » – seront regroupés dans une même cohorte. Chaque navigateur partagera un identifiant de cohorte, indiquant le groupe d'appartenance, avec les sites web et les annonceurs. D'après la proposition, chaque cohorte devrait contenir au moins plusieurs milliers d'utilisatrices et utilisateurs (ce n'est cependant pas une garantie).

Si cela vous semble complexe, imaginez ceci : votre identifiant AFC sera comme un court résumé de votre activité récente sur le Web.

La démonstration de faisabilité de Google utilisait les noms de domaines des sites visités comme base pour grouper les personnes. Puis un algorithme du nom de SimHash permettait de créer les groupes. Il peut tourner localement sur la machine de tout un chacun, il n'y a donc pas besoin d'un serveur central qui collecte les données comportementales. Toutefois, un serveur administrateur central pourrait jouer un rôle dans la mise en œuvre des garanties de confidentialité. Afin d'éviter qu'une cohorte soit trop petite (c'est à dire trop caractéristique), Google propose qu'un acteur central puisse compter le nombre de personnes dans chaque cohorte. Si certaines sont trop petites, elles pourront être fusionnées avec d'autres cohortes similaires, jusqu'à ce qu'elles représentent suffisamment d'utilisateurs.

Pour que l'AFC soit utile aux publicitaires, une cohorte d'utilisateurs ou utilisatrices devra forcément dévoiler des informations sur leur comportement.

Selon la proposition formulée par Google, la plupart des spécifications sont déjà à l'étude. Le projet de spécification prévoit que l'identification d'une cohorte sera accessible via JavaScript, mais on ne peut pas savoir clairement s'il y aura des restrictions, qui pourra y accéder ou si l'identifiant de l'utilisateur sera partagé par d'autres moyens. L'AFC pourra

constituer des groupes basés sur l'URL ou le contenu d'une page au lieu des noms domaines ; également utiliser une synergie de « système apprentissage » (comme le sous-entend l'appellation AFC) afin de créer des regroupements plutôt que de se baser sur l'algorithme de SimHash. Le nombre total de cohortes possibles n'est pas clair non plus. Le test de Google utilise une cohorte d'utilisateurs avec des identifiants sur 8 bits, ce qui suppose qu'il devrait y avoir une limite de 256 cohortes possibles. En pratique, ce nombre pourrait être bien supérieur ; c'est ce que suggère la documentation en évoquant une « cohorte d'utilisateurs en 16 bits comprenant 4 caractères hexadécimaux ». Plus les cohortes seront nombreuses, plus elles seront spécialisées – plus les identifiants de cohortes seront longs, plus les annonceurs en apprendront sur les intérêts de chaque utilisatrice et auront de facilité pour cibler leur empreinte numérique.

Mais si l'un des points est déjà clair c'est le facteur temps. Les cohortes AFC seront réévaluées chaque semaine, en utilisant chaque fois les données recueillies lors de la navigation de la semaine précédente.

Ceci rendra les cohortes d'utilisateurs moins utiles comme identifiants à long terme, mais les rendra plus intrusives sur les comportements des utilisatrices dans la durée.

De nouveaux problèmes pour la vie privée.

L'AFC fait partie d'un ensemble qui a pour but d'apporter de la publicité ciblée dans un futur où la vie privée serait préservée. Cependant la conception même de cette technique implique le partage de nouvelles données avec les annonceurs. Sans surprise, ceci crée et ajoute des risques concernant la donnée privée.

Le Traçage par reconnaissance d'ID.

Le premier enjeu, c'est le pistage des navigateurs, une pratique qui consiste à collecter de multiples données

distinctes afin de créer un identifiant unique, personnalisé et stable lié à un navigateur en particulier. Le projet Cover Your Tracks (Masquer Vos Traces) de l'Electronic Frontier Foundation (EFF) montre comment ce procédé fonctionne : pour faire simple, plus votre navigateur paraît se comporter ou agir différemment des autres, plus il est facile d'en identifier l'empreinte unique.

Google a promis que la grande majorité des cohortes AFC comprendrait chacune des milliers d'utilisatrices, et qu'ainsi on ne pourra vous distinguer parmi le millier de personnes qui vous ressemblent. Mais rien que cela offre un avantage évident aux pisteurs. Si un pistage commence avec votre cohorte, il doit seulement identifier votre navigateur parmi le millier d'autres (au lieu de plusieurs centaines de millions). En termes de théorie de l'information, les cohortes contiendront quelques bits d'entropie jusqu'à 8, selon la preuve de faisabilité. Cette information est d'autant plus éloquente sachant qu'il est peu probable qu'elle soit corrélée avec d'autres informations exposées par le navigateur. Cela va rendre la tâche encore plus facile aux traqueurs de rassembler une empreinte unique pour les utilisateurs de l'AFC.

Google a admis que c'est un défi et s'est engagé à le résoudre dans le cadre d'un plan plus large, le « Budget vie privée » qui doit régler le problème du pistage par l'empreinte numérique sur le long terme. Un but admirable en soi, et une proposition qui va dans le bon sens ! Mais selon la Foire Aux Questions, le plan est « une première proposition, et n'a pas encore d'implémentation dans un navigateur ». En attendant, Google a commencé à tester l'AFC dès ce mois de mars.

Le pistage par l'empreinte numérique est évidemment difficile à arrêter. Des navigateurs comme Safari et Tor se sont engagés dans une longue bataille d'usure contre les pisteurs, sacrifiant une grande partie de leurs fonctionnalités afin de réduire la surface des attaques par traçage. La limitation du pistage implique généralement des coupes ou des restrictions

sur certaines sources d'entropie non nécessaires. Il ne faut pas que Google crée de *nouveaux* risques d'être tracé tant que les problèmes liés aux risques existants subsistent.

L'exposition croisée

Un second problème est moins facile à expliquer : la technologie va partager de nouvelles données personnelles avec des pisteurs qui peuvent déjà identifier des utilisatrices. Pour que l'AFC soit utile aux publicitaires, une cohorte devra nécessairement dévoiler des informations comportementales.

La page Github du projet aborde ce sujet de manière très directe :

Cette API démocratise les accès à certaines informations sur l'historique de navigation général des personnes (et, de fait, leurs intérêts principaux) à tous les sites qui le demandent... Les sites qui connaissent les Données à Caractère Personnel (c'est-à-dire lorsqu'une personne s'authentifie avec son adresse courriel) peuvent enregistrer et exposer leur cohorte. Cela implique que les informations sur les intérêts individuels peuvent éventuellement être rendues publiques.

Comme décrit précédemment, les cohortes AFC ne devraient pas fonctionner en tant qu'identifiant intrinsèque. Cependant, toute entreprise capable d'identifier un utilisateur d'une manière ou d'une autre – par exemple en offrant les services « identifiez-vous via Google » à différents sites internet – seront à même de relier les informations qu'elle apprend de l'AFC avec le profil de l'utilisateur.

Deux catégories d'informations peuvent alors être exposées :

1. Des informations précises sur l'historique de navigation. Les pisteurs pourraient mettre en place une rétro-ingénierie sur l'algorithme d'assignation des cohortes pour savoir si une

utilisatrice qui appartient à une cohorte spécifique a *probablement* ou *certainement* visité des sites spécifiques.

2. Des informations générales relatives à la démographie ou aux centres d'intérêts. Par exemple, une cohorte particulière pourrait sur-représenter des personnes jeunes, de sexe féminin, ou noires ; une autre cohorte des personnes d'âge moyen votant Républicain ; une troisième des jeunes LGBTQ+, etc.

Cela veut dire que chaque site que vous visitez se fera une bonne idée de quel type de personne vous êtes dès le premier contact avec ledit site, sans avoir à se donner la peine de vous suivre sur le Net. De plus, comme votre cohorte sera mise à jour au cours du temps, les sites sur lesquels vous êtes identifié·e·s pourront aussi suivre l'évolution des changements de votre navigation. Souvenez-vous, une cohorte AFC n'est ni plus ni moins qu'un résumé de votre activité récente de navigation.

Vous devriez pourtant avoir le droit de présenter différents aspects de votre identité dans différents contextes. Si vous visitez un site pour des informations médicales, vous pourriez lui faire confiance en ce qui concerne les informations sur votre santé, mais il n'y a pas de raison qu'il ait besoin de connaître votre orientation politique. De même, si vous visitez un site de vente au détail, ce dernier n'a pas besoin de savoir si vous vous êtes renseigné·e récemment sur un traitement pour la dépression. L'AFC érode la séparation des contextes et, au contraire, présente le même résumé comportemental à tous ceux avec qui vous interagissez.

Au-delà de la vie privée

L'AFC est conçu pour éviter une menace spécifique : le profilage individuel qui est permis aujourd'hui par le croisement des identifiants contextuels. Le but de l'AFC et des autres propositions est d'éviter de laisser aux pisteurs l'accès à des informations qu'ils peuvent lier à des gens en

particulier. Alors que, comme nous l'avons montré, cette technologie pourrait aider les pisteurs dans de nombreux contextes. Mais même si Google est capable de retravailler sur ses conceptions et de prévenir certains risques, les maux de la publicité ciblée ne se limitent pas aux violations de la vie privée. L'objectif même de l'AFC est en contradiction avec d'autres libertés individuelles.

Pouvoir cibler c'est pouvoir discriminer. Par définition, les publicités ciblées autorisent les annonceurs à atteindre certains types de personnes et à en exclure d'autres. Un système de ciblage peut être utilisé pour décider qui pourra consulter une annonce d'emploi ou une offre pour un prêt immobilier aussi facilement qu'il le fait pour promouvoir des chaussures.

Au fur et à mesure des années, les rouages de la publicité ciblée ont souvent été utilisés pour l'exploitation, la discrimination et pour nuire. La capacité de cibler des personnes en fonction de l'ethnie, la religion, le genre, l'âge ou la compétence permet des publicités discriminatoires pour l'emploi, le logement ou le crédit. Le ciblage qui repose sur l'historique du crédit – ou des caractéristiques systématiquement associées – permet de la publicité prédatrice pour des prêts à haut taux d'intérêt. Le ciblage basé sur la démographie, la localisation et l'affiliation politique aide les fournisseurs de désinformation politique et la suppression des votants. Tous les types de ciblage comportementaux augmentent les risques d'abus de confiance.

Au lieu de réinventer la roue du pistage, nous devrions imaginer un monde sans les nombreux problèmes posés par les publicités ciblées.

Google, Facebook et beaucoup d'autres plateformes sont en train de restreindre certains usages sur de leur système de ciblage. Par exemple, Google propose de limiter la capacité des annonceurs de cibler les utilisatrices selon des «

catégories de centres d'intérêt à caractère sensible ». Cependant, régulièrement ces tentatives tournent court, les grands acteurs pouvant facilement trouver des compromis et contourner les « plateformes à usage restreint » grâce à certaines manières de cibler ou certains types de publicité.

Même un imaginant un contrôle total sur quelles informations peuvent être utilisées pour cibler quelles personnes, les plateformes demeurent trop souvent incapables d'empêcher les usages abusifs de leur technologie. Or l'AFC utilisera un algorithme non supervisé pour créer ses propres cohortes. Autrement dit, personne n'aura un contrôle direct sur la façon dont les gens seront regroupés.

Idéalement (selon les annonceurs), les cohortes permettront de créer des regroupements qui pourront avoir des comportements et des intérêts communs. Mais le comportement en ligne est déterminé par toutes sortes de critères sensibles : démographiques comme le genre, le groupe ethnique, l'âge ou le revenu ; selon les traits de personnalités du « Big 5 » ; et même la santé mentale. Ceci laisse à penser que l'AFC regroupera aussi des utilisateurs parmi n'importe quel de ces axes.

L'AFC pourra aussi directement rediriger l'utilisatrice et sa cohorte vers des sites internet qui traitent l'abus de substances prohibées, de difficultés financières ou encore d'assistance aux victimes d'un traumatisme.

Google a proposé de superviser les résultats du système pour analyser toute corrélation avec ces catégories sensibles. Si l'on découvre qu'une cohorte spécifique est étroitement liée à un groupe spécifique protégé, le serveur d'administration pourra choisir de nouveaux paramètres pour l'algorithme et demander aux navigateurs des utilisateurs concernés de se constituer en un autre groupe.

Cette solution semble à la fois orwellienne et digne de Sisyphe. Pour pouvoir analyser comment les groupes AFC seront associés à des catégories sensibles, Google devra mener des

enquêtes gigantesques en utilisant des données sur les utilisatrices : genre, race, religion, âge, état de santé, situation financière. Chaque fois que Google trouvera qu'une cohorte est associée trop fortement à l'un de ces facteurs, il faudra reconfigurer l'ensemble de l'algorithme et essayer à nouveau, en espérant qu'aucune autre « catégorie sensible » ne sera impliquée dans la nouvelle version. Il s'agit d'une variante bien plus compliquée d'un problème que Google s'efforce déjà de tenter de résoudre, avec de fréquents échecs.

Dans un monde numérique doté de l'AFC, il pourrait être plus difficile de cibler directement les utilisatrices en fonction de leur âge, genre ou revenu. Mais ce ne serait pas impossible. Certains pisteurs qui ont accès à des informations secondaires sur les utilisateurs seront capables de déduire ce que signifient les groupes AFC, c'est-à-dire quelles catégories de personnes appartiennent à une cohorte, à force d'observations et d'expérimentations. Ceux qui seront déterminés à le faire auront la possibilité de la discrimination. Pire, les plateformes auront encore plus de mal qu'aujourd'hui à contrôler ces pratiques. Les publicitaires animés de mauvaises intentions pourront être dans un déni crédible puisque, après tout, ils ne cibleront pas directement des catégories protégées, ils viseront seulement les individus en fonction de leur comportement. Et l'ensemble du système sera encore plus opaque pour les utilisatrices et les régulateurs.

Cookies, c'est fini, 🎵
dire que c'était la source
de mes premiers revenus.



et maintenant, un peu de FLoC and Roll !



Avec Google les instruments changent, mais c'est toujours la même musique...

Google, ne faites pas ça, s'il vous plaît

Nous nous sommes déjà prononcés sur l'AFC et son lot de propositions initiales lorsque tout cela a été présenté pour la première fois, en décrivant l'AFC comme une technologie « contraire à la vie privée ». Nous avons espéré que les processus de vérification des standards mettraient l'accent sur les défauts de base de l'AFC et inciteraient Google à renoncer à son projet. Bien entendu, plusieurs problèmes soulevés sur leur GitHub officiel exposaient exactement les mêmes préoccupations que les nôtres. Et pourtant, Google a poursuivi le développement de son système, sans pratiquement rien changer de fondamental. Ils ont commencé à déployer leur discours sur l'AFC auprès des publicitaires, en vantant le remplacement du ciblage basé sur les cookies par l'AFC « avec une efficacité de 95 % ». Et à partir de la version 89 de Chrome, depuis le 2 mars, la technologie est déployée pour un galop d'essai. Une petite fraction d'utilisateurs de Chrome – ce qui fait tout de même plusieurs millions – a été assignée aux tests de cette nouvelle technologie.

Ne vous y trompez pas, si Google poursuit encore son projet d'implémenter l'AFC dans Chrome, il donnera probablement à

chacun les « options » nécessaires. Le système laissera probablement le choix par défaut aux publicitaires qui en tireront bénéfice, mais sera imposé par défaut aux utilisateurs qui en seront affectés. Google se glorifiera certainement de ce pas en avant vers « la transparence et le contrôle par l'utilisateur », en sachant pertinemment que l'énorme majorité de ceux-ci ne comprendront pas comment fonctionne l'AFC et que très peu d'entre eux choisiront de désactiver cette fonctionnalité. L'entreprise se félicitera elle-même d'avoir initié une nouvelle ère de confidentialité sur le Web, débarrassée des vilains cookies tiers, cette même technologie que Google a contribué à développer bien au-delà de sa date limite, engrangeant des milliards de dollars au passage.

Ce n'est pas une fatalité. Les parties les plus importantes du bac-à-sable de la confidentialité comme l'abandon des identificateurs tiers ou la lutte contre le pistage des empreintes numériques vont réellement améliorer le Web. Google peut choisir de démanteler le vieil échafaudage de surveillance sans le remplacer par une nouveauté nuisible.

Nous rejetons vigoureusement le devenir de l'AFC. Ce n'est pas le monde que nous voulons, ni celui que méritent les utilisatrices. Google a besoin de tirer des leçons pertinentes de l'époque du pistage par des tiers et doit concevoir son navigateur pour l'activité de ses utilisateurs et utilisatrices, pas pour les publicitaires.

Remarque : nous avons contacté Google pour vérifier certains éléments exposés dans ce billet ainsi que pour demander davantage d'informations sur le test initial en cours. Nous n'avons reçu aucune réponse à ce jour.