GOUVERNEMENT

Liberté
Égalité
Fraternité

Pôle d'expertise de la régulation numérique

Éclairage sur…

Mars 2022

#03

Privacy Sandbox : une collection d'outils pour une publicité en ligne exempte de cookies tiers

Aujourd'hui, l'écosystème de la publicité en ligne repose en grande partie sur l'exploitation de cookies tiers afin d'améliorer la pertinence des annonces présentées aux utilisateurs. Google a décidé de supprimer à l'horizon 2023 ces cookies tiers de son navigateur Chrome, leader mondial du marché. Il propose une alternative technologique baptisée *Privacy Sandbox* : une collection d'outils présentée par Google comme à la fois plus respectueuse de la vie privée et viable économiquement.

En cours de développement, la *Privacy Sandbox* est amenée à évoluer régulièrement. Ce numéro d'Éclairage sur… fait le point sur les technologies du dispositif connues à sa date de parution et en passe d'être déployées. À cette date, l'ensemble des cas d'usage permis par les cookies tiers ne sont pas encore reproductibles par le projet *Privacy Sandbox*. Ce dernier n'a pas démontré sa capacité à permettre aux acteurs de monétiser la publicité aussi bien qu'ils le faisaient avec les cookies tiers. Dans ce contexte, un avantage de fond de Google semble se dessiner : de plus en plus souvent en position de *first-party*, il serait ainsi moins affecté que ses concurrents par une efficacité réduite du projet *Privacy Sandbox*.

QUELQUES CONCEPTS CLÉS POUR COMPRENDRE L'INITIATIVE *PRIVACY SANDBOX*

La publicité en ligne : adapter le message au contexte et aux personnes

Selon la CNIL[1], la publicité ciblée (ou personnalisée) est une « technique publicitaire qui vise à identifier les personnes individuellement afin de leur diffuser des messages publicitaires spécifiques en fonction de caractéristiques individuelles ». L'initiative *Privacy Sandbox* aborde le cadre plus large de la publicité programmatique. Celle-ci inclut la publicité dite contextuelle, quand elle se base sur le contenu affiché sur la même page, ou la publicité par ciblage comportemental, lorsque la segmentation se fait en fonction de l'historique de navigation d'un utilisateur. Être capable de cibler un utilisateur (en identifiant, par exemple, ses centres d'intérêts) lors de sa navigation sert à l'ensemble de l'écosystème de la publicité en ligne pour acheter, suivre, analyser, mesurer, contrôler et adapter les campagnes publicitaires en temps réel.

Publicité programmatique
Permet de planifier l'achat d'espaces publicitaires de manière automatique selon des critères prédéfinis (prix, caractéristiques de l'audience, heure de la journée …).

Le marché mondial de la publicité en ligne représentait 378 milliards[2] de dollars en 2020 et finance en grande partie le web dit gratuit (applications, ou sites, par exemple certains blogs et articles journalistiques). La publicité ciblée constitue une part conséquente des revenus générés. Pour fonctionner elle s'appuie sur une collecte et une exploitation massive des données des utilisateurs.

Google, qui tire la majorité de ses revenus de la publicité en ligne[3] dont il est un acteur majeur sinon dominant, indique devoir désormais concilier la protection de la vie privée de ses utilisateurs avec l'impact économique que pourrait avoir la suppression des cookies tiers sur l'écosystème. Ainsi le projet *Privacy Sandbox* a pour ambition de permettre à la publicité programmatique de continuer de fonctionner dans un monde sans cookies tiers, tout en préservant la confidentialité des internautes vis-à-vis des tiers.

Les cookies : une technologie socle pour la publicité en ligne

First-Party
On désigne par *first-party* un acteur qui interagit directement avec son audience. Ainsi, les données récoltées par Cdiscount lorsqu'un utilisateur navigue sur cdiscount.com sont dites *first-party*.

Pour réaliser le ciblage des internautes, **l'écosystème de la publicité en ligne utilise aujourd'hui massivement les technologies fondées sur les cookies.** Ces fichiers contenant du texte, enregistrés sur le navigateur Internet et associés à un nom de domaine[4], permettent de stocker des informations sur la navigation et les actions de l'internaute. Deux types de cookies sont à l'œuvre : les cookies *first-party* et les cookies tiers.

---

[1] CNIL, Définitions - Publicité ciblée (https://www.cnil.fr/fr/definition/publicite-ciblee).

[2] Statista, *Digital advertising spending worldwide from 2019 to 2024 (in billion U.S. dollars),* mai 2021 (https://www.statista.com/statistics/237974/online-advertising-spending-worldwide/).

[3] Plus de 80% de son chiffre d'affaires soit environ 147 milliards de dollars en 2020
Alphabet Inc., *Alphabet Announces Fourth Quarter and Fiscal Year 2020 Results*, février 2021 (https://abc.xyz/investor/static/pdf/2020Q4_alphabet_earnings_release.pdf).

[4] Par exemple : lemonde.fr, pour plus de détails, consulter Afnic.fr (https://www.afnic.fr/noms-de-domaine/tout-savoir/).

Déposés par le site internet visité, **les cookies *first-party*** ont principalement vocation à permettre le fonctionnement du site et à améliorer l'expérience de l'utilisateur (par exemple, garder une connexion active, mémoriser un panier sur une place de marché, effectuer des mesures d'audience, etc). En outre, le site peut conserver des informations sur l'utilisateur et sa navigation, même si celui-ci ne s'est pas authentifié.

Au-delà de cette interaction directe entre un site et un visiteur, une page web peut afficher des contenus provenant d'autres sites internet : vidéos, images, podcasts, polices de caractère, etc. Ces sites internet tiers sont alors en mesure de déposer et de lire leurs propres cookies, **dits cookies « tiers »**, sur l'ordinateur de l'internaute via son navigateur.

Illustration : si un utilisateur navigue d'un site A à un blog B proposant tous deux des vidéos hébergées sur un site tiers C, le site C pourra alors déposer et lire ses propres cookies sur le navigateur de l'internaute se laissant pour lui-même une trace du parcours de l'utilisateur, d'abord lors de la visite du site A puis du blog B. **Ce mécanisme a permis l'émergence d'entreprises spécialisées dans le suivi des utilisateurs. Dans le cadre de partenariats, certaines d'entre elles sont présentes sur pratiquement tous les sites visités par les internautes et sont donc au fait de l'ensemble de leur navigation**. Cette navigation est analysée pour en déduire des préférences et des profils d'utilisateurs qui permettent ainsi de faire du ciblage publicitaire. Les sites directement visités par les internautes, comme A ou B, acceptent de laisser des sites comme C suivre leurs utilisateurs, la contrepartie étant de pouvoir mieux monétiser l'affichage de publicités ciblées sur leurs pages. **Le suivi des utilisateurs via les cookies tiers permet ainsi aux entreprises de construire des profils, en utilisant un ensemble large de données contenant aussi des données à caractère personnel. Même si ce suivi requiert le consentement des utilisateurs, il n'en reste pas moins peu transparent et difficile à maîtriser par ces derniers, tout comme le choix des données transmises.**

Vers la fin des cookies tiers ?

Pour des raisons de protection de la vie privée, Mozilla et Apple ont progressivement restreint les cookies tiers dans leurs navigateurs Firefox et Safari[5]. Ainsi, en 2019, Firefox a bloqué les cookies tiers identifiés comme liés au suivi, puis segmenté en 2021 ceux qui n'étaient pas encore bloqués, en fonction du site à l'origine de leur création. Safari, quant à lui, a initié en 2017 un blocage de plus en plus restrictif des cookies tiers avec la première mouture de l'*Intelligent Tracking Prevention* de webkit.

[5] Mozilla, Firefox bloque désormais par défaut les cookies tiers de pistage et les mineurs de cryptomonnaies, 3 septembre 2019 (https://blog.mozilla.org/press-fr/2019/09/03/firefox-bloque-desormais-par-defaut-les-cookies-tiers-de-pistage-et-les-mineurs-de-cryptomonnaies/). Sabharwal, C. Safari ITP: *Intelligent Tracking Prevention Version 1.0 to 2.3*. *Adpushup*, 10 Septembre 2020 (https://www.adpushup.com/blog/safari-itp-intelligent-tracking-prevention/).
Huang, T., Hofmann J., Edelstein A. *Firefox 86 Introduces Total Cookie Protection*. Mozilla, 23 février 2021 (https://blog.mozilla.org/security/2021/02/23/total-cookie-protection/).

Dans les deux cas, cela se traduit par une baisse potentielle de revenus pour les acteurs du ciblage publicitaire, mais également pour les sites visités[6].

De son côté, Google, leader du marché avec son navigateur Internet Chrome (plus de 70% de parts de marché en 2021[7]) a lancé en 2019 une initiative open source au sein du forum de standardisation *World Wide Web Consortium* (W3C). **Baptisée *Privacy Sandbox*[8], cette initiative a pour objectif annoncé d'intégrer, d'ici 2023, au sein de Chrome, une solution répondant à l'ensemble des usages publicitaires couverts par les cookies tiers tout en protégeant mieux les données personnelles (cf. *infra*) et en limitant le suivi des utilisateurs**.

Privacy Sandbox est une série d'outils (en grande partie des APIs), encore au stade de proposition ou d'expérimentation, qui devrait servir à remplacer à terme l'ensemble des fonctionnalités permises par les cookies tiers.

W3C
World Wide Web Consortium, organisme de standardisation à but non lucratif. Travaille au développement des standards du web tel que HTML, CSS …

API
Application Programming Interface ou interface de programmation applicative interface logicielle qui permet de « connecter » un logiciel ou un service à un autre logiciel ou service afin d'échanger des données et des fonctionnalités.

PRIVACY SANDBOX : DYNAMIQUE D'UN PROJET SUIVI DE PRÈS PAR LES ACTEURS

C'est en réponse à l'enjeu grandissant de protection de la vie privée et à différents scandales[9] que Mozilla et Apple ont limité l'accès aux données de leurs utilisateurs, au détriment des revenus de l'industrie[10] publicitaire.

À son tour, Google a initié en 2019 ses travaux sur les outils de la *Privacy Sandbox*. Si l'objectif initial de mise en œuvre était fixé à 2022, celui-ci a été repoussé à au moins 2023. Chaque outil du projet *Privacy Sandbox* est décrit publiquement, et la liste de ces outils est disponible sur le site de présentation du projet[11]. Tous ces outils appartiennent à différentes catégories telles que « Afficher du contenu et des publicités pertinentes », « Mesurer l'efficacité des campagnes » ou en encore « Combattre le spam et la fraude »[12]. Une partie des travaux autour de ces outils est examinée au sein de groupes de travail du W3C, dont Google est un membre influent.

---

[6] Hern, A. *No tracking, no revenue: Apple's privacy feature costs ad companies millions*. The Guardian, 9 janvier 2018 (https://www.theguardian.com/technology/2018/jan/09/apple-tracking-block-costs-advertising-companies-millions-dollars-criteo-web-browser-safari).

[7] CNIL, Alternatives aux cookies tiers : quelles conséquences en matière de consentement, 13 octobre 2001 (https://www.cnil.fr/fr/alternatives-aux-cookies-tiers-quelles-consequences-en-matiere-de-consentement).

[8] Voir le site de la *Privacy Sandbox* (https://privacysandbox.com/).

[9] *Ad Industry Accused Of 'Massive' Privacy Breach*, Forbes, 18 janvier 2019 (https://www.forbes.com/sites/emmawoollacott/2019/01/28/ad-industry-accused-of-massive-privacy-breach/).

[10] Ibid, 6.

Note : Sur cette question, des plaintes contre Apple ont été déposées, soutenant que celui-ci s'abriterait derrière un motif de protection de la vie privée des utilisateurs pour avantager ses propres services, notamment de publicité ciblée au sein de son écosystème (en cours d'instruction au fond à l'Autorité de la concurrence à ce jour).

[11] Ibid, 8.

[12] *Privacy Sandbox Overview*, en date du 31/03/2022 (https://privacysandbox.com/open-web/#how-works-on-web-hero)

Ainsi, la première moitié de l'année 2020 voit un florilège d'outils de la *Privacy Sandbox* rapidement passer la phase d'incubation (cf. *infra* pour leur description). Certains arrivent en phase d'expérimentation dès 2021. Dès ces premières expérimentations, des doutes sont exprimés[13] sur les conséquences des technologies promues par Google et leurs effets sur la concurrence. En juin 2021, une annonce de procédure négociée est publiée par l'autorité britannique de la concurrence, la *Competition and Markets Authority* (CMA[14]), craignant notamment que la *Privacy Sandbox* ne renforce la position de Google dans l'écosystème de la publicité[15]. Google prend de premiers engagements et promet plus de transparence autour de l'initiative. Ainsi, un calendrier prévisionnel officiel est finalement publié pour les différents outils[16]. En novembre 2021, après une consultation par la CMA de différents acteurs du secteur, Google renforce ses engagements, sans répondre complètement aux inquiétudes émises par l'autorité britannique[17].

En parallèle, Criteo, spécialiste français de la publicité ciblée, met en doute la capacité de la *Privacy Sandbox* à permettre aux acteurs de la publicité en ligne de continuer leurs activités avec les mêmes fonctionnalités que celles permises par les cookies tiers[18]. Criteo a réalisé des expérimentations internes sur un outil d'affichage de publicité ciblée de la *Privacy Sandbox* : *FLoC* [19], et a conclu, entre autres, qu'il est difficile à l'heure actuelle d'assurer que la *Privacy Sandbox* générera autant de revenus pour les acteurs que les anciennes méthodes.

CONSÉQUENCES SUR LA VALEUR POUR LA PUBLICITÉ EN LIGNE

Différentes critiques sont adressées au système envisagé par Google, dont il convient de rappeler qu'il n'est pas finalisé à ce stade et qu'il pourrait être modifié avant son éventuel lancement pour corriger les défauts identifiés.

L'ensemble des cas d'usage permis par les technologies de cookies ne sont pas encore reproductibles avec les outils de la *Privacy Sandbox*. Il est par exemple impossible d'utiliser des vidéos publicitaires[20] via FLEDGE, outil de ciblage publicitaire.

---

[13] WIRED, *Antitrust and Privacy are on a collision course*, 12 avril 2021 (https://www.wired.com/story/antitrust-privacy-on-collision-course)

[14] Competion and Market Authority (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/992975/Notice_of_intention_to_accept_binding_commitments_offered_by_Google_publication.pdf)

[15] Voir par exemple le point 5.3 du rapport de la CMA (Ibid, 14).

[16] Ibid, 12.

[17] Voir par exemple les points 2.3 et 2.4 de l'analyse et publication des nouveaux engagements de Google par la CMA (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1036204/211126_FINAL_modification_notice.pdf).

[18] La série « *FLoC Origin Trial* » sur le blog Criteo Tech, Medium dont le premier (https://medium.com/criteo-engineering/floc-origin-trial-what-we-observed-3f7e8f209b82).

[19] *FLoC* sur Github (https://github.com/WICG/floc).

[20] Video advertising on the web #29, Github (https://github.com/WICG/turtledove/issues/29).

Monétisation
Les contenus gratuits en ligne (presses, vidéos…) sont souvent financés via la publicité en ligne. On parle de monétisation du contenu lorsque l'éditeur se rémunère sans faire payer à l'utilisateur mais en lui proposant des publicités.

La capacité de la solution *Privacy Sandbox* à monétiser aussi bien que les technologies de cookies n'est pas démontrée. En effet, s'il y a moins de suivi individuel et moins de fonctionnalités disponibles qu'avec le recours aux cookies tiers, il est possible qu'il y ait moins d'opportunités pour créer de la valeur. Or, si *Privacy Sandbox* ne parvenait pas à monétiser aussi bien que les technologies de cookies tiers, cela n'aurait de conséquences que pour les acteurs « tiers ». En effet les acteurs s'appuyant sur les cookies *first-party*, notamment dans des environnements exigeant une authentification de l'utilisateur ne sont pas affectés par ce changement technologique. Ainsi le changement n'entrave pas la capacité des entreprises à collecter des données via des cookies *first-party*. Les protagonistes pouvant se passer des cookies tiers, typiquement via un environnement logué, ne seront en rien lésés et connaîtront même une concurrence moindre.

Dans ce contexte, un avantage de fond de Google semble se dessiner : de plus en plus souvent en position de *first-party*, il serait donc moins affecté que ses concurrents par une efficacité réduite de la *Privacy Sandbox*. En effet, Google accroît progressivement l'hébergement de contenus directement sur ses propres serveurs au motif notamment d'offrir des temps de réponse plus rapides à ses utilisateurs. Par exemple, en intégrant directement des réponses à ses pages de résultats (météo, horaires de cinéma, etc) ou en favorisant les contenus au format AMP (*Accelerated Mobile Pages*), que Google met en cache sur ses serveurs. Héberger ces contenus lui permet de faire un suivi en tant que *first-party*, sans perte d'information, de toutes les activités des internautes qui restent ainsi sur ses sites et serveurs. À l'inverse, les auteurs de contenus diffusés via AMP, puisqu'ils n'hébergent plus ces contenus désormais accessibles via le cache du moteur de recherche, deviennent tierces parties et dépendent de Google pour avoir les informations les plus détaillées concernant la lecture de leurs propres contenus.

Enfin, quand bien même la monétisation via la *Privacy Sandbox* serait équivalente à celle permise par les cookies tiers et que l'ensemble des acteurs réaliseraient le saut technologique vers la *Privacy Sandbox* sans aucun retard ni coût d'apprentissage, il subsisterait un risque économique pour les sites et les acteurs de la publicité ciblée lié à une **modification de la répartition de la valeur entre acteurs**. En effet, les mécanismes de la *Privacy Sandbox* ayant pour but d'empêcher la collecte et le croisement d'informations provenant de différents sites, certains acteurs comme les *Data Management Platform* ou DMP[21] verront probablement une partie de leur activité réduite.

[21] CNIL, Data management platform (DMP) ou « plateforme de gestion des données » (https://www.cnil.fr/fr/definition/data-management-platform-dmp-ou-plateforme-de-gestion-des-donnees)

PRÉSENTATION DE COMPOSANTS TECHNIQUES DE LA *PRIVACY SANDBOX*

Afin de rendre plus concrets les changements qu'impliquent la *Privacy Sandbox*, cette partie présente quatre outils du projet :

- *FLoC*, *Topics* et *FLEDGE* qui répondent au besoin de présenter des publicités pertinentes à chaque utilisateur ;
- *Trust Token* API qui est une solution technique permettant de lutter contre la fraude.

A noter que d'autres outils de la Privacy Sandbox abordent des axes différents tels que le *reporting* des campagnes publicitaires et la lutte contre le *tracking* entre les sites.

Tracking
Pratique visant à associer à l'utilisateur un identifiant unique afin de suivre sa navigation au fil des sites.

Des solutions de ciblage par intérêt : de *FLoC* à *Topics*

La publicité basée sur les centres d'intérêts (*Interest Based Advertising* ou IBA) repose sur la collecte de données sur des domaines web détenus ou exploités par différentes entités. La publicité présentée à l'internaute est adaptée à ses préférences ou intérêts connus ou déduits de son historique de navigation par exemple.

Un premier outil de ciblage a été présenté par Google : *FLoC*[22]. L'approche proposée était de regrouper les internautes en fonction de la similarité de leurs historiques de navigation dans des cohortes. Chaque cohorte correspond à un mélange d'intérêts déduits des noms de domaine visités. Pour s'assurer du respect de la vie privée, la décision d'affectation dans une cohorte était prise périodiquement par le navigateur en local.

Fingerprinting
Technique visant à identifier un utilisateur de façon unique grâce aux informations communiquées par son matériel. Par exemple, son adresse IP, la version de son navigateur …

Du point de vue de la vie privée, le numéro unique de cohorte a été très critiqué[23] car il aurait pu servir de point d'entrée pour identifier l'internaute via des mécanismes de *fingerprinting*. De plus, la lecture de cet identifiant sur le navigateur nécessitant un consentement dans le cadre du RGPD, Google n'a pu conduire sa phase de test dans les pays où ce règlement s'applique, entravant la capacité des acteurs à étudier la solution.

Pour les acteurs de la publicité en ligne, le numéro de cohorte n'apportait pas directement d'information sémantique et nécessitait donc des investissements pour pouvoir la rattacher à des intérêts généraux. Cela aurait pu avoir pour effet d'augmenter l'asymétrie entre petits et grands acteurs. **Paradoxalement, comme n'importe quel acteur pouvait accéder à l'identifiant de cohorte, ce système tendait à répartir gratuitement l'information de navigation** entre tous les acteurs au détriment des acteurs qui créent cette valeur : les éditeurs de contenus.

[22] Ibid. 19.

[23] *Electronic Frontier Foundation*, *Google FLoC Is a Terrible Idea*, 3 mars 2021 (https://www.eff.org/fr/deeplinks/2021/03/googles-floc-terrible-idea)

À la suite de ces nombreuses critiques, une nouvelle façon d'aborder cette problématique est proposée en janvier 2022 : l'outil *Topics*[24]. Le principe : chaque semaine, le navigateur associe à l'utilisateur ses 5 centres d'intérêt (*topics*) préférés, sur la base des sites qu'il visite. Ce top est conservé 3 semaines par le navigateur, qui peut alors enregistrer au maximum 15 centres d'intérêt.

À ce jour, la liste complète des centres d'intérêt est déterminée pour ne pas être susceptible de révéler d'information sensible. Vouée à évoluer, elle comprend aujourd'hui 349[25] occurrences. Le navigateur prédit directement les intérêts d'une navigation par un modèle de *machine learning* qui analyse les noms de domaines visités. Chaque site est ainsi associé à un ou plusieurs centres d'intérêt, ces derniers sont comptabilisés à chaque visite de l'utilisateur. À la fin de la semaine, les 5 centres les plus rencontrés constituent les centres « préférés » d'un utilisateur.

Les acteurs n'ont accès qu'à une liste restreinte de centres d'intérêt associés à leur site ou ceux de leurs partenaires, leur liste observable de centres d'intérêt. Ils ne peuvent recevoir l'information qu'un utilisateur (anonyme) possède un centre d'intérêt particulier qu'à la double condition d'avoir observé cet utilisateur sur un site, et que le centre d'intérêt présenté par l'API fasse partie de leur liste observable. Ce mécanisme imite celui des cookies tiers : il est en effet nécessaire de nouer des partenariats avec des sites pour obtenir de l'information, ou bien posséder soi-même plusieurs sites.

En pratique, lorsqu'un internaute visitera un site, son navigateur choisira aléatoirement 3 centres d'intérêt parmi les 15 centres enregistrés. Chaque acteur présent sur ce site (le propriétaire du site et ses partenaires), pourra alors recevoir pendant une semaine les centres d'intérêts qui auront été sélectionnés, et qui seront présents dans sa liste observable.

La notion de centres d'intérêt telle que proposé par Topics est bien plus directement exploitable que le numéro de cohorte sans signification prévu par FLoC. Néanmoins, le mécanisme de « top 5 », cumulé à un élargissement probable des centres d'intérêt possibles, laisse craindre une prépondérance des sujets très généraux au détriment des sujets spécifiques, limitant ainsi la personnalisation des publicités et la valeur ajoutée qu'elles génèrent pour tout l'écosystème. Contrairement à ce qui était prévu pour *FLoC*, *Topics* sera activé uniquement sur les sites qui en feront la demande, en revanche, il le sera par défaut coté utilisateur.

Une solution de (re)ciblage publicitaire : *FLEDGE*

FLEDGE[26] doit permettre la diffusion d'une publicité auprès d'un internaute « potentiellement intéressé » qui a déjà interagi avec le site

Re-ciblage
Stratégie marketing permettant à un annonceur de cibler spécifiquement ses visiteurs qui n'ont pas encore réalisé l'action désirée (typiquement un achat).

[24] Sur Github (https://github.com/jkarlin/topics).

[25] Sur Github (https://github.com/jkarlin/topics/blob/main/taxonomy_v1.md).

[26] *FLEDGE* sur Github (https://github.com/WICG/turtledove/blob/main/FLEDGE.md#summary).

de l'annonceur ou le réseau publicitaire que ce dernier utilise. Plus exactement, *FLEDGE* propose d'une part, une solution pour créer et mettre à jour des **groupes d'intérêt** et d'autre part, un **mécanisme d'enchère** sur le navigateur de l'internaute. Il faut bien distinguer les « groupes d'intérêt » de *FLEDGE* des « centres d'intérêts » de *Topics*. Dans *FLEDGE* un groupe d'intérêt est juste un groupement d'utilisateurs partageant une action commune sur un site ou groupe de sites (par exemple une visite, un achat, un clic).

Dans sa démarche consultative, Google s'est appuyé sur les retours de nombreux acteurs[27] sur son outil initial, *TURTLEDOVE*, pour développer son remplaçant, *FLEDGE*.

FLEDGE est construit pour que le navigateur opère les enchères et contienne les informations sur l'appartenance de l'internaute aux groupes d'intérêt. Le site sur lequel la publicité s'affiche ne devrait donc pas être en mesure de connaître le groupe remportant l'enchère. De plus, l'annonceur doit uniquement baser son enchère sur le groupe d'intérêt et ne doit pas pouvoir identifier les personnes qui le composent. Dans le cas contraire, il serait en mesure de croiser l'intérêt de l'utilisateur avec d'autres informations.

FLEDGE est au cœur de *Privacy Sandbox* car il permet de gérer concrètement certains cas d'usage habituels de la publicité en ligne. Quelques exemples :

- un site de vente en ligne peut facilement effectuer du re-ciblage publicitaire. Pour cela, il ajoute ses visiteurs dans un groupe d'intérêt.
- un forum de voiture peut qualifier son audience comme étant intéressée par l'automobile. Il permet à des tiers de cibler son groupe d'intérêt associé à l'automobile.
- un éditeur peut déléguer la création de groupes d'intérêt à d'autres acteurs spécialisés dans la qualification d'utilisateurs. À première vue, ce mécanisme renforce la position des sites éditeurs en leur assurant le contrôle de leurs données.

Néanmoins, un acteur en partenariat avec plusieurs éditeurs pourrait créer un groupe sur plusieurs sites à la fois, lui donnant une position plus favorable pour ajouter des utilisateurs dans les groupes. Il est possible qu'in fine, l'avantage revienne aux intermédiaires - par exemple les *Sell Side Platform* - en leur donnant une position plus favorable pour créer et peupler des groupes.

Toutefois, *FLEDGE* est toujours en phase de construction. Son défi est de pouvoir réaliser une enchère publicitaire au sein du navigateur en limitant l'accès des acteurs aux données utilisateurs. Afin de réaliser l'enchère, le navigateur doit échanger certaines informations avec les acteurs publicitaires[28]. Sans mécanisme approprié, ces échanges sont

[27] RTB House, NextRoll, Magnite, Criteo, et Google Ads team.

[28] Par exemple, le navigateur peut demander à l'annonceur associé à un groupe d'intérêt quel est le budget restant pour cette campagne.

l'occasion pour les acteurs de collecter ou de déduire des informations sur les utilisateurs. *FLEDGE*, à terme, devrait utiliser un mécanisme de serveur décentralisé[29] qui n'est pas encore défini. La première expérimentation de l'outil utilisera donc les serveurs des acteurs impliqués[30], ce qui limite considérablement les garanties recherchées en termes de protection de la vie privée. Sans plus d'informations, à l'heure actuelle, il est difficile de pronostiquer comment ce défi technique sera relevé.

Transférer l'information sur l'authenticité d'un internaute : *Trust Token* API

Trust Token API est un projet d'outil visant à partager la confiance qu'établit un site en un utilisateur, comme déterminer son authenticité, par exemple en s'assurant qu'il n'est pas un *bot*.

Bot
Programme informatique automatisé ayant pour but de simuler le comportement d'une personne humaine et d'effectuer des tâches. En publicité des bots peuvent être utilisé pour générer du faux trafic sur un site, générant ainsi des coûts publicitaires factices pour les annonceurs.

L'avantage théorique est double :

- pour les sites, avoir des garanties d'authenticité du trafic à moindre coût ;
- pour les utilisateurs, diminuer le nombre de *captchas* ou moyens *ad-hoc* de vérifier qu'ils sont légitimes.

Ce projet d'apparence légitime et sans lien direct avec le ciblage ou la publicité, risque néanmoins de renforcer la capacité de certains acteurs, comme Google, à surveiller l'évolution du trafic de nombreux sites web.

Dans une vision simplifiée, **cette méthode distingue deux rôles : les sites émetteurs, qui vérifient la légitimité des utilisateurs et partagent l'information, et les sites récepteurs de cette confiance.** La confiance serait diffusée par le biais d'un « jeton » anonyme fourni à l'internaute par un site émetteur. Lorsque l'internaute visite un autre site - appelé récepteur - alors il demande au site émetteur de valider l'authenticité du jeton (pour éviter une accumulation de ces jetons ou leur utilisation frauduleuse).

Le rôle d'émetteur présente de nombreuses difficultés dans le paramétrage (pour limiter les risques d'attaque et pour améliorer la détection des *bots*), ainsi que des coûts (de développement, de serveurs) possiblement importants. Les sites récepteurs ont intérêt à choisir les jetons des sites émetteurs qui identifient le mieux le trafic légitime, induisant une compétition entre les sites émetteurs. Cependant, aucun mécanisme de rémunération pour ce type d'acteurs ne semble prévu. Ainsi, un ingénieur de Facebook s'interroge[31] sur l'intérêt qu'aurait un site à se charger du rôle d'émetteur, mais également sur les risques portant sur le respect de la vie privée des utilisateurs, par exemple parce qu'un utilisateur présentant un jeton d'authentification émis par une

[29] Le modèle « *Trusted-Server* » prévoit un serveur détenant les informations nécessaires à l'enchère, ne gardant aucune trace des échanges avec les utilisateurs.

[30] « *Bring Your Own Server* » : durant l'expérimentation, chaque acteur sera libre d'échanger avec son propre serveur.

[31] Sur Github (https://github.com/WICG/trust-token-api/issues/28)

plateforme donnée serait forcément un utilisateur de cette même plateforme.

L'ensemble des contraintes fait que ce rôle d'émetteur de jetons pourrait être peu attractif et n'intéresser qu'un nombre limité d'acteurs. Il pourrait paradoxalement se révéler particulièrement intéressant pour certains acteurs, dont Google. D'une part, l'émission de jetons et la remontée en temps réel des demandes de confiance pourrait offrir aux acteurs concernés une connaissance anonymisée mais très fine des sites visités par les internautes ainsi que de leur trafic. Cette connaissance pourrait être à même de constituer un avantage sur des marchés connexes (utile au classement pour le référencement, information stratégique sur l'audience de concurrents, etc.). D'autre part, il ne faut pas exclure à terme que ce service présenté comme gratuit devienne payant, soit pour les utilisateurs, soit pour les sites qui ont besoin de savoir que les visiteurs sont authentiques. S'il devient impossible ou trop complexe de certifier l'authenticité d'un utilisateur sans ces *trust tokens*, délivrés par un nombre très restreint d'acteurs, une monétisation n'est ainsi pas à exclure à moyen ou long terme.

UNE EXTENSION DU PROJET : LA *PRIVACY SANDBOX* SUR ANDROID

En février 2022, Google annonce vouloir étendre le projet de la *Privacy Sandbox* à l'univers mobile. L'objectif est similaire au projet web : permettre de fournir aux applications des revenus via la publicité tout en limitant le *tracking* et notamment l'usage du Google *Advertising ID*, un identifiant unique[32] à vocation publicitaire.

Google se donne 2 ans pour adapter certains des outils de la *Privacy Sandbox* web et en développer de nouveaux[33], les participations à cette initiative se feront directement dans la section développeur du site d'Android[34].

La *Privacy Sandbox* sur mobile va constituer un projet à part entière et nécessitera une analyse approfondie au fil de ses avancements.

[32] Google Play, Identifiant publicitaire (https://support.google.com/googleplay/android-developer/answer/6048248)

[33] Par exemple SDK Runtime sur *Android* Developers (https://developer.android.com/design-for-safety/ads/sdk-runtime).

[34] *Privacy Sandbox on Android* (https://developer.android.com/design-for-safety/ads).

Dépôt légal : Mars 2022
ISSN (en ligne): 2824-8201
contact.peren@finances.gouv.fr

Service à compétence nationale, le Pôle d'expertise de la régulation numérique (PEReN) fournit, aux services de l'État et autorités administratives intervenant dans la régulation des plateformes numériques, une expertise et une assistance technique dans les domaines du traitement des données, des data sciences et des procédés algorithmiques. Il s'investit également dans des projets de recherche en science des données à caractère exploratoire ou scientifique.

PEReN – 120 rue de Bercy, 75572 Paris Cedex 12