GOUVERNEMENT

Liberté
Égalité
Fraternité

Pôle d'expertise de la régulation numérique

Éclairage sur…

Mai 2024

08

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

Version actualisée

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 internautes. Après avoir envisagé 2024, c'est finalement en 2025 que Google supprimera 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 test ou d'adoption progressive par les acteurs de la publicité en ligne, la *Privacy Sandbox* avait, lors de son lancement, fait l'objet d'une première exploration par le PEReN dans son n°3 d'« Éclairage sur… » (Mars 2022). Ce nouvel opus consacré à la solution Google passe en revue les évolutions observées depuis.

À ce jour, la *Privacy Sandbox* n'a pas encore démontré sa capacité à permettre aux éditeurs de monétiser la publicité autant qu'ils le faisaient avec les *cookies* tiers. Dans ce contexte de limitation du partage d'informations publicitaires entre acteurs, un avantage de fond semble se dessiner en faveur des plus grands acteurs qui peuvent récolter de manière autonome une grande quantité de données. Tout comme cela avait été observé en 2022, ils garderaient un avantage relatif et seraient moins affectés que leurs concurrents par une efficacité réduite de la suite d'outils *Privacy Sandbox*.

SOMMAIRE

L'ESSENTIEL EN UNE PAGE

Pour des raisons de protection de la vie privée, Mozilla et Apple ont progressivement supprimé ou restreint l'usage des *cookies* tiers dans leurs navigateurs. Souhaitant suivre ses concurrents, Google a lancé **le projet *Privacy Sandbox*. L'objectif : intégrer, au sein de Chrome, une suite d'outils répondant à certains des usages publicitaires couverts par les *cookies* tiers, tout en limitant le suivi des internautes**.

Le projet *Privacy Sandbox* ambitionne de « Lutter contre le spam et la fraude sur le web », « Mesurer les annonces numériques » ou encore « Afficher des publicités et des contenus pertinents ». Cette dernière catégorie regroupe les outils du projet les plus cruciaux pour l'écosystème :

- l'API *Topics* : intégrée au navigateur, elle associe aux internautes des centres d'intérêts à partir des sites visités ;
- l'API *Protected Audience* : elle permet aux annonceurs de regrouper les internautes au sein de groupes d'intérêts qu'il est ensuite possible de cibler lors d'enchères sécurisées.

À ce jour, la *Privacy Sandbox* ne fait pas consensus auprès des navigateurs concurrents de Chrome. Ces derniers s'opposent en grande partie à l'intégration, au sein du navigateur, d'outils purement publicitaires.

La capacité de la *Privacy Sandbox* à monétiser aussi bien que les technologies de *cookies* tiers n'est pas démontrée à ce jour. Contrairement aux éditeurs reposant sur une authentification (notamment les réseaux sociaux), les sites web accessibles librement (dont la plupart des sites de presse) risquent de voir la valeur de leurs encarts publicitaires baisser si le ciblage proposé par la *Privacy Sandbox* est moins performant que celui reposant sur les *cookies* tiers.

Peu après le début des expérimentations, l'autorité britannique de la concurrence (CMA) annonce, en juin 2021, une procédure négociée afin de s'assurer que la *Privacy Sandbox* ne renforcera pas la position de Google. Face notamment aux inquiétudes de l'autorité et des acteurs, l'objectif initial de retrait complet des cookies tiers a été repoussé progressivement de 2022 à 2025.

Depuis l'annonce de la disparition des *cookies* tiers sur Chrome, des acteurs ont travaillé à des solutions alternatives. Certaines intègrent des mécanismes fort de protection de la vie privée quand d'autres permettent de continuer à suivre de près les parcours des internautes sans recours aux *cookies* tiers. Ces nouveaux outils pourraient venir compléter ou concurrencer les outils de la *Privacy Sandbox*.

Cette année 2024 devrait révéler les évolutions du marché de la publicité en ligne, entre adoption plus large de la *Privacy Sandbox* et recours aux solutions alternatives.

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

L'écosystème de la publicité en ligne : un panorama

La publicité en ligne consiste en général à vendre un espace publicitaire sur un site web dont les contenus attirent des internautes (presse en ligne, forums, réseaux sociaux) à une entreprise qui souhaite faire la publicité de ses produits ou services. Ces sites web sont nommés **éditeurs** car ils éditent des contenus. On appelle **annonceurs** les entreprises qui souhaitent bénéficier de la visibilité des contenus des éditeurs afin de faire leur publicité. Sur le marché de la publicité en ligne, les éditeurs représentent ainsi **l'offre** et les annonceurs, **la demande**.

La valeur de la publicité affichée dépend de l'intérêt que les internautes qui la visionnent lui portent. Une publicité hors de leurs centres d'intérêt ou invasive a peu de chance de déclencher un achat ou d'avoir un effet positif, et génère donc moins de revenus. Ainsi, l'objectif des éditeurs et des annonceurs est souvent aligné : afficher les publicités les plus pertinentes, qui ont le plus de valeur. Des campagnes efficaces auront pour effet d'augmenter à la fois les revenus de l'annonceur et la valeur de l'emplacement publicitaire fourni par l'éditeur.

Cet objectif d'optimisation de la pertinence a historiquement entraîné l'émergence :

- **d'intermédiaires** spécialisés, au premier rang desquels figurent les **plateformes d'enchères (*ad exchange*).** Celles-ci ont vocation à maximiser la valeur des encarts publicitaires par une meilleure ouverture de l'offre et de la demande.
- de la publicité ciblée : une « technique publicitaire qui vise à identifier les personnes individuellement afin de leur diffuser des messages publicitaires spécifiques en fonction de caractéristiques individuelles » (définition de la CNIL[1])

Le paramétrage des campagnes publicitaires via les plateformes d'enchères nécessite des compétences techniques spécifiques souvent coûteuses à acquérir pour les éditeurs et les annonceurs qui se tournent alors vers :

- les **plateformes de vente ou SSP (*supply-side platforms*).** Elles permettent aux éditeurs de déléguer la vente de leurs encarts. En pratique, les SSP et les *ad exchange* sont parfois confondus.
- les **plateformes d'achat ou DSP (*demand-side platforms*).** Elles permettent aux annonceurs de déléguer l'optimisation des enchères publicitaires et le paramétrage des encarts ou audiences sur lesquels enchérir.

Le marché mondial de la publicité en ligne représentait 624.6 milliards[2] de dollars en 2023 et finance en grande partie le web dit gratuit (applications, sites, blogs, articles journalistiques…). La publicité ciblée

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

[2] Statista, digital advertising – worldwide : https://fr.statista.com/outlook/dmo/digital-advertising/worldwide#ad-spending

constitue une part conséquente des revenus générés. Pour fonctionner, elle s'appuie sur une collecte et une exploitation massive des données des internautes.

Au sein du marché de la publicité en ligne, Google propose des services sur l'ensemble de la chaîne de valeur :

- Les internautes utilisent majoritairement le navigateur **Chrome, leader du marché**[3] pour accéder aux sites des éditeurs[4] ;
- **Google Ad Manager (**fusion d'AdX – plateforme d'enchères de Google – et de [2]DoubleClick for publishers**)**, l'une des SSP de Google, concentre la majorité des parts de marché des SSP ;
- **DV360 et Google Ads**, les DSP de Google, détiennent aussi de larges parts de marché.

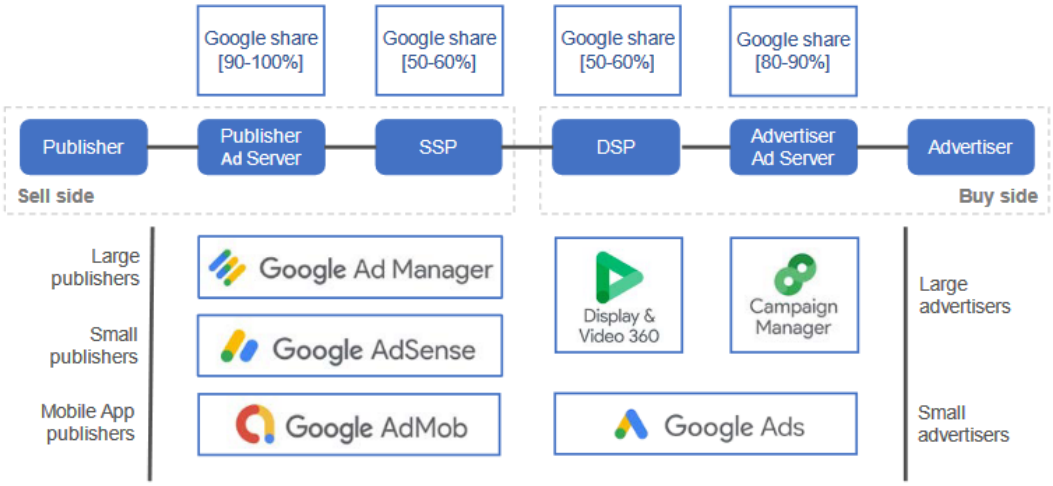

Figure 1 : La place de Google dans l'intermédiation publicitaire (source : Competition and Market Authority, autorité britannique de la concurrence)[5].

Les *cookies* : une technologie socle pour la publicité en ligne

Pour réaliser le ciblage des internautes, **l'écosystème de la publicité en ligne utilise encore 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[6], permettent de stocker des informations sur la navigation et les actions de l'internaute.

[3] Chrome représente environ 60% de parts de marché en France en 2023, Statcounter, *Browser Market Share France :* https://gs.statcounter.com/browser-market-share/all/france.

[4] Android, un système d'exploitation pour smartphones détenu par Google, est aussi majoritaire dans son domaine. Des encarts publicitaires sont également mis en vente au sein des applications (dont l'application Chrome, installée par défaut). Les applications ont sur mobile un rôle similaire à celui des éditeurs sur le web.

[5] Figure 5.15, p. 271 : https://assets.publishing.service.gov.uk/media/5efc57ed3a6f4023d242ed56/Final_report_1_July_2020_.pdf

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

Deux types de *cookies* sont à l'œuvre : les *cookies first-party*[7] et les *cookies* tiers.

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'internaute (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'internaute et sa navigation, même si celui-ci ne s'est pas authentifié.

Au-delà de cette interaction directe entre un site et un internaute, 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.

Voici une illustration par l'exemple. Un internaute navigue du site de presse « Figamondération.fr » vers le site de commerce « Cdiscoin.com ». Ces deux sites hébergeant chacun du contenu du site tiers « MyFavoriteAd.com », le propriétaire de « MyFavoriteAd.com » pourra alors déposer et lire ses propres *cookies* sur le navigateur de l'internaute. Il conserve ainsi une trace du parcours de l'internaute sur le site « Figamondération.com » puis sur le site « Cdiscoin.com ». Cette navigation est analysée pour en déduire des préférences et des profils d'internautes qui permettent ainsi de faire du ciblage publicitaire. Les sites directement visités par les internautes, comme « Figamondération.fr » ou « Cdiscoin.com », acceptent de laisser des sites comme « MyFavoriteAd.com » suivre leurs utilisateurs. En contrepartie, ces sites sont mieux rémunérés via l'affichage de publicités ciblées sur leurs pages.

Ce mécanisme a permis l'émergence d'entreprises spécialisées dans le suivi des internautes. 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 d'une grande partie de leur navigation. Le suivi des internautes 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 reste peu transparent et en pratique difficile à appréhender pour ces derniers : quelles données sont transmises ? À qui ? Pour quelles fins ?

[7] 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*.

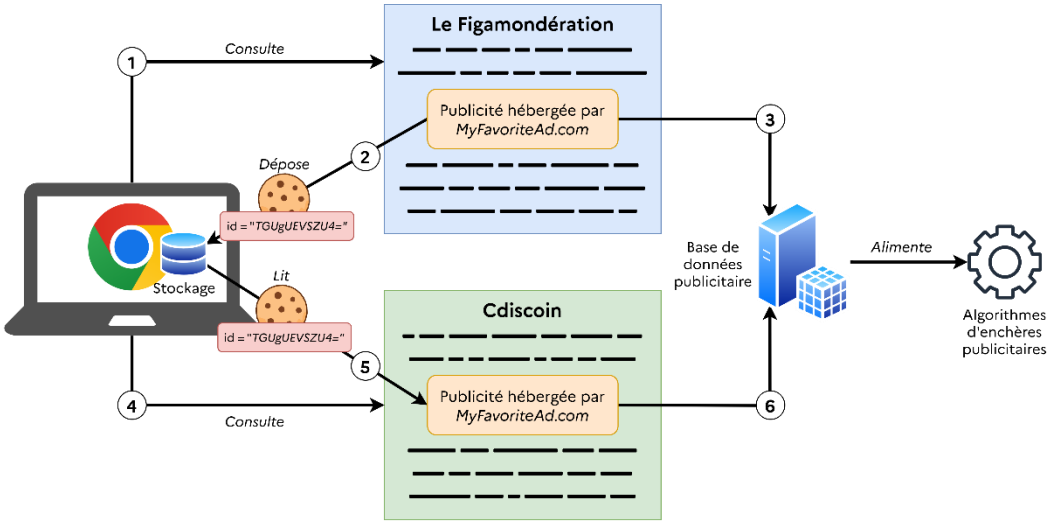

Figure 2 : Illustration du suivi publicitaire par l'intermédiaire de cookies tiers

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[8]. 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*. 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[9].
De son côté, Google, a lancé en 2019 une initiative *open source* au sein du forum de standardisation *World Wide Web Consortium* (W3C)[10]. **Baptisée *Privacy Sandbox*[11], cette initiative a pour objectif d'intégrer, au sein de Chrome et d'autres navigateurs si les développeurs le souhaitent, une suite d'outils (majoritairement des APIs[12]) répondant à certains des usages publicitaires couverts par les *cookies* tiers tout en protégeant mieux les données personnelles et en limitant le suivi des internautes**.

[8] 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/

[9] 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

[10] Le *World Wide Web Consortium (W3C)* est un organisme de standardisation à but non lucratif qui travaille au développement des standards du web tel que HTML, CSS ...

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

[12] Une API (*Application Programming Interface* ou interface de programmation applicative) est une 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.

Destinée à remplacer les fonctionnalités permises par les *cookies* tiers, la *Privacy Sandbox* est entrée, fin 2023, dans une phase de mise à disposition dans Chrome (dite « *General availability* [13]»). Éditeurs, annonceurs et intermédiaires de l'écosystème de la publicité en ligne peuvent désormais se saisir des outils pour évaluer et se préparer au retrait total des *cookies* tiers prévu en 2025[14].

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*, première initiative qui a été remplacée par l'API *Topics,* et l'API *Protected Audience* (ou PAAPI) qui répondent au besoin de présenter des publicités pertinentes à chaque internaute ;
- L'API *Private State Tokens,* 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*[15] entre les sites, quelles que soient les méthodes utilisées.

Des solutions de ciblage par intérêt : de *FLoC* à l'API *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 en 2019 : *FLoC*[16]. L'approche proposée était de regrouper les internautes dans des cohortes en fonction de la similarité de leurs historiques de navigation. 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.

Du point de vue de la vie privée, le numéro unique de cohorte a été très critiqué[17] car il aurait pu servir de point d'entrée pour identifier l'internaute via des mécanismes de *fingerprinting*[18]. De plus, la lecture de cet identifiant sur le navigateur nécessitant un consentement dans le cadre du Règlement Général sur la Protection des Données (RGPD),

[13] https://privacysandbox.com/open-web/#the-privacy-sandbox-timeline

[14] Une section dédiée « *Third party cookies and testing* » de la timeline renseigne sur les étapes de disparition des *cookies*.

[15] Le *tracking* associe à l'utilisateur un identifiant unique afin de suivre sa navigation au fil des sites.

[16] *FLoC* sur Github : https://github.com/WICG/floc

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

[18] Le *fingerprinting* est une technique visant à identifier un utilisateur de façon unique grâce aux informations communiquées par son matériel (adresse IP, version du navigateur…)

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 être relié à des intérêts généraux. Cela aurait pu avoir pour effet d'augmenter l'asymétrie d'informations disponibles 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 de ceux qui créent cette valeur, les éditeurs de contenus.

À la suite de ces critiques, une nouvelle façon d'aborder cette problématique est proposée en janvier 2022 : l'API *Topics*[19], ci-après *Topics*. 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. Cette association est déterminée en local, c'est-à-dire sur le terminal de l'utilisateur. Ce top est conservé 3 semaines par le navigateur, qui peut alors enregistrer au maximum 15 centres d'intérêt.

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 469[20] items. Les centres d'intérêt sont répartis sous forme d'arborescence de familles : par exemple « Art & Divertissement » est un parent de « Art & Divertissement / Films / Films de Comédie », et réciproquement « Films » est un enfant de « Art & Divertissement ». Depuis fin 2023, les grandes familles de centres d'intérêt ont été réparties en deux groupes de valeur, « Standard » et « Haute » afin de prioriser les familles à forte valeur ajoutée.

Pour associer à un site, un ou plusieurs centres d'intérêt, le navigateur s'appuie sur un modèle de *machine learning* embarqué. Conscient des limites actuelles de ce modèle, Google a choisi de prédéfinir les centres d'intérêt pour 50 000 sites[21].

Chaque site est ainsi associé à un ou plusieurs centres d'intérêt 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, en priorisant les intérêts des familles à valeur « Haute ».

Les acteurs n'ont ainsi accès qu'à une liste restreinte de centres d'intérêt associés à leur site ou ceux de leurs partenaires, qui constituent 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 condition d'avoir observé cet utilisateur sur un site associé à ce centre d'intérêt ou un centre d'intérêt enfant. Ce mécanisme imite celui des *cookies* tiers : il est en effet nécessaire de nouer des partenariats avec

[19] Topics sur Github : https://github.com/jkarlin/topics

[20] Sur Github : https://github.com/patcg-individual-drafts/topics/blob/main/taxonomy_v2.md

[21] « includes a larger override list—50k rather than 10k websites », Google au 11 janvier 2024 : https://developers.google.com/privacy-sandbox/relevance/topics/latest?hl=en#classifier

des sites visités par un utilisateur pour obtenir de l'information sur cet utilisateur, 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 (l'éditeur 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 graduel 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.

L'activation de *Topics* pour un internaute est pour le moment soumise à une fenêtre de consentement spécifique. Elle s'affiche à l'installation de Chrome ou à l'utilisation pour les utilisateurs historiques.

Une solution de (re)ciblage publicitaire[22] : l'API *Protected Audience*

L'API *Protected Audience*[23], ci-après « *PAAPI »,* doit permettre la diffusion d'une publicité auprès d'un internaute « potentiellement intéressé » qui a déjà interagi avec le site de l'annonceur ou le réseau publicitaire que ce dernier utilise. Plus exactement, *PAAPI* propose d'une part, une solution pour créer et mettre à jour des **groupes d'intérêt,** et d'autre part, un **mécanisme d'enchères sur le navigateur de l'internaute**. Il faut bien distinguer les « groupes d'intérêt » de *PAAPI* des « centres d'intérêts » de *Topics*. Dans *PAAPI,* un groupe d'intérêt représente un groupe 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[24] sur son outil initial, *TURTLEDOVE*, pour développer *PAAPI* (connue pendant un temps sous le nom de *FLEDGE)*.

PAAPI est construite pour que le navigateur opère, en local, 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.

---

[22] Le re-ciblage publicitaire est une stratégie marketing permettant à un annonceur de cibler spécifiquement des visiteurs déjà rencontrés. En règle générale, il s'agit d'un utilisateur qui n'a pas encore réalisé l'action désirée (typiquement un achat).

[23] L'API *Protected Audience* sur Github : https://github.com/WICG/turtledove/blob/main/FLEDGE.md#summary

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

PAAPI est au cœur de *Privacy Sandbox* car elle 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.

Un acteur publicitaire partenaire de plusieurs éditeurs pourrait créer un groupe rassemblant des utilisateurs provenant de plusieurs sites. Ainsi, les acteurs publicitaires présents sur un grand nombre de sites auront une position plus favorable pour créer des groupes et les peupler.

Le défi de *PAAPI* est de permettre des enchères publicitaires ciblées au sein du navigateur tout en limitant l'accès aux données utilisateur. En effet, pour réaliser l'enchère, le navigateur doit échanger certaines informations avec les intermédiaires publicitaires[25]. Sans mécanisme de protection approprié, ces échanges sont l'occasion pour les acteurs de collecter ou de déduire des informations sur les utilisateurs. *PAAPI* devrait, après le retrait des *cookies* tiers, rendre obligatoire l'utilisation d'un mécanisme de serveur décentralisé[26] dit *« Trusted Execution Environment »* (TEE*).* Dans l'intervalle, les acteurs publicitaires peuvent temporairement utiliser leur propre serveur (modèle « *Bring Your Own Server* »[27]), avec des garanties de sécurité moindres. À noter, aujourd'hui, Google Cloud Platform et Amazon Web Services[28] sont les seuls acteurs à fournir un service de TEE dans le cadre de la *Privacy Sandbox*. Au vu des exigences cibles[29], il semble que peu d'acteurs seront en mesure d'offrir le service.

PAAPI permet de valoriser les groupes via l'organisation d'enchères. En plus des enchères contextuelles basées sur les données *first-party* dont disposent les éditeurs ou les SSP, ces derniers ont la possibilité d'initier une ou plusieurs enchères via *PAAPI*. L'organisation de plusieurs enchères prévoit un système à deux niveaux avec une enchère dite « pilote » (*top-*

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

[26] Le modèle « *Trusted-Server* » prévoit un serveur détenant les informations nécessaires à l'enchère, ne gardant aucune trace des échanges avec les utilisateurs. Voir Github – *Trusted-Execution Environment* pour plus d'informations : https://github.com/privacysandbox/fledge-docs/blob/main/trusted_services_overview.md#trusted-execution-environment

[27] Voir la partie « *fetching real-time data* », Github : https://github.com/WICG/turtledove/blob/main/FLEDGE.md#31-fetching-real-time-data-from-a-trusted-server

[28] Relevé en mai 2024, disponibilité des plateformes de Cloud : https://developer.chrome.com/en/docs/privacy-sandbox/aggregation-service/#availability

[29] *Public Cloud TEE Requirements, Github* : https://github.com/privacysandbox/protected-auction-services-docs/blob/main/public_cloud_tees.md#security-and-trust-of-the-csp

level) composée de plusieurs autres enchères dites « composantes » (*component auctions*). Contrairement aux enchères composantes, l'enchère pilote a accès à la valeur finale des enchères contextuelles[30]. Alléguant la protection des intérêts des éditeurs, Google Ad Manager (GAM) a annoncé ne participer à aucune enchère dans le cadre de *PAAPI* s'il n'est pas responsable de l'enchère « pilote »[31]. Étant donné la position très majoritaire de GAM, cela risque d'entraver la capacité d'autres SSP à effectuer des enchères en tant que « pilote », les éditeurs qu'ils représenteront ne souhaitant pas se couper des offres de Google et renoncer à une partie importante du marché. De surcroît, chaque organisateur d'enchère peut envoyer des données contextuelles différentes à chaque potentiel acheteur, à la discrétion de l'organisateur de l'enchère. L'implémentation technique de ce partage d'information via le SSP devra faire l'objet d'une attention particulière, afin de garantir que l'éditeur garde effectivement le contrôle sur les informations communiquées aux acheteurs potentiels de ses espaces publicitaires.

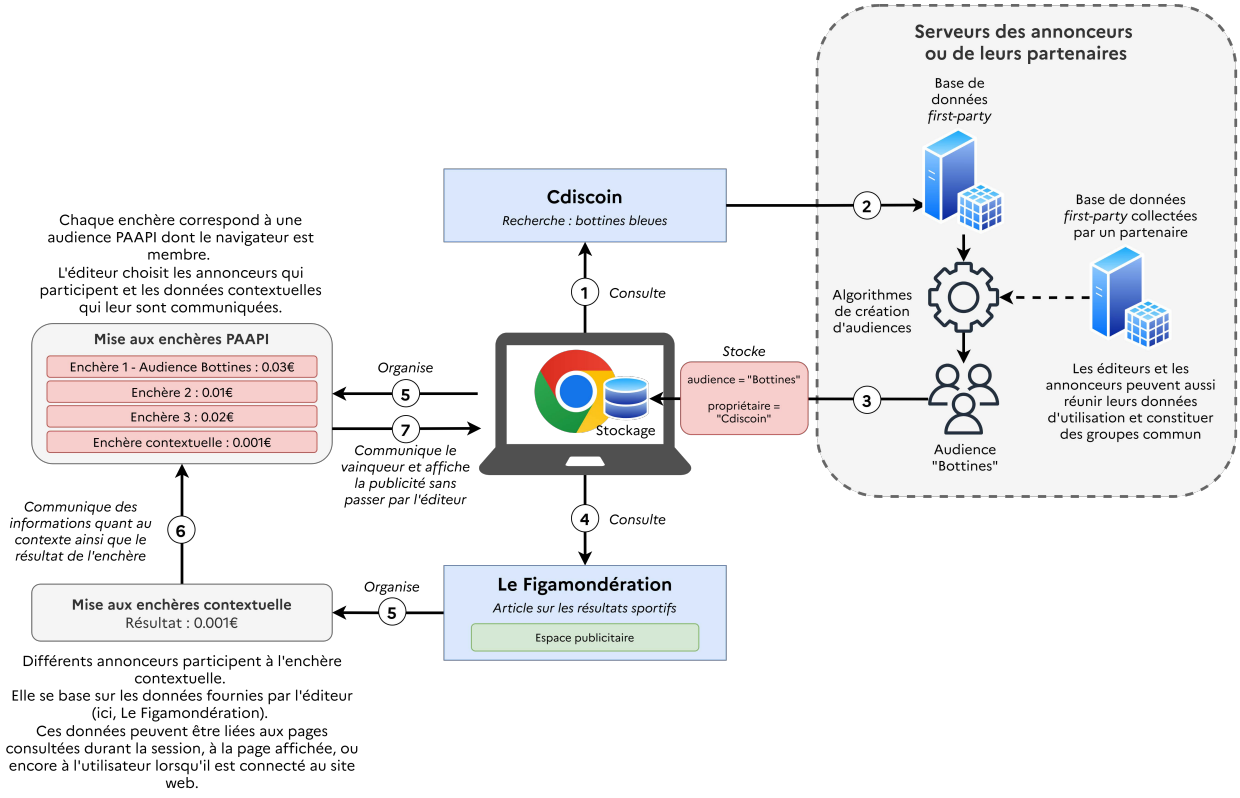

Figure 3 : Illustration du mécanisme des enchères organisées par l'API Protected Audience (PAAPI) au sein du navigateur Chrome. *Chaque enchère prend la forme de code informatique exécuté en local par le navigateur. En cas de victoire d'une l'enchère, le navigateur récupère la publicité sur un serveur et l'affiche.*

[30] Dans le cas d'une enchère avec composantes, toutes les enchères (*bids)* sont récoltées par les enchères (*auctions)* composantes et l'enchère pilote choisit parmi elles (voir 2.1 : https://github.com/WICG/turtledove/blob/main/FLEDGE.md#21-initiating-an-on-device-auction)

[31] Ce comportement est mentionné dans le dernier rapport trimestriel 2023 de la CMA relatif à la Privacy Sandbox : https://assets.publishing.service.gov.uk/media/65ba2a504ec51d000dc9f1f5/CMA_Q4_2023_update_report_on_implementation_of_the_Privacy_Sandbox_commitments_PDFA_1.pdf. Voir la partie « Protected Audience API », entrée « Top-Level Auctions » à l'adresse suivante pour la réponse de Google : https://developers.google.com/privacy-sandbox/overview/feedback/report-2023-q4

Transférer l'information sur l'authenticité d'un internaute : l'API *Private State Tokens*

L'API *Private State Tokens* 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*[32].

L'avantage théorique est double :

- pour les sites, avoir des garanties d'authenticité du trafic sans avoir recours aux *cookies* tiers ;
- pour les utilisateurs, limiter le recours à des outils de vérification basés sur l'utilisation de données personnelles, et potentiellement diminuer le nombre de *captchas* ou moyens *ad-hoc* de vérifier qu'ils sont légitimes.

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 – celui-ci demande au site émetteur de valider l'authenticité du jeton. Le site émetteur renvoie alors à l'internaute un enregistrement d'utilisation (*redemption records)* qui a valeur d'attestation de son authenticité.

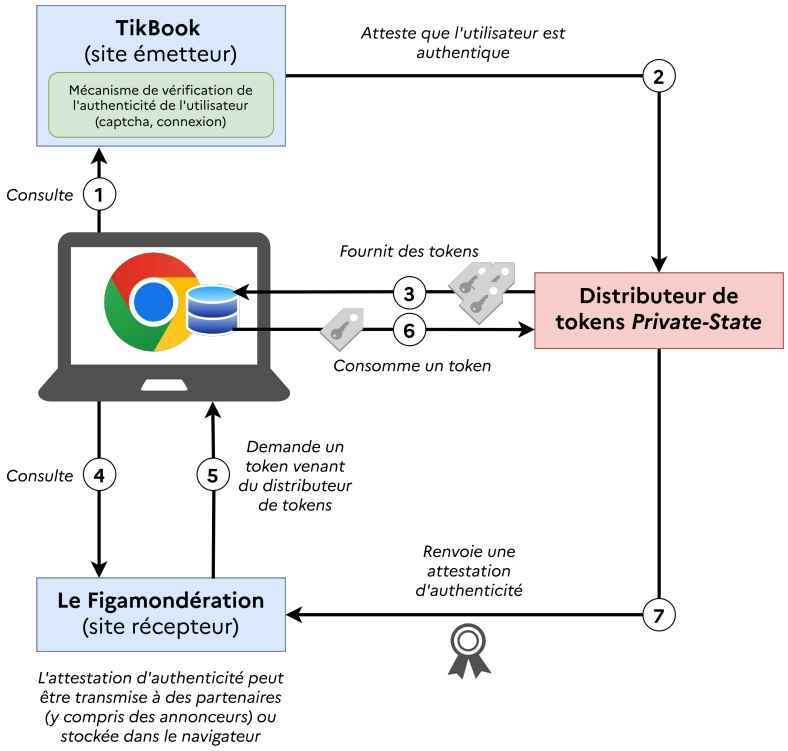

Figure 4 : Illustration du transfert de l'information sur l'authenticité d'un internaute par l'API Private State Tokens au sein du navigateur Chrome

[32] Un *bot* est un 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.

Le rôle d'émetteur présente de nombreuses difficultés. Un processus d'enregistrement auprès de Google est nécessaire tous les 6 mois[33] et la mise en place du système implique des coûts (de développement, de serveurs) potentiellement importants. À cette date, seulement une poignée d'acteurs ont réalisé la procédure d'enregistrement[34]. Pour le moment, ce sont principalement des entreprises proposant des services de Captcha.

Les sites récepteurs ayant intérêt à choisir les jetons des sites émetteurs qui identifient le mieux le trafic légitime, ceci induirait une concurrence entre les sites émetteurs. Cependant, aucun mécanisme de rémunération associé à la fourniture de ce service ne semble prévu. Dès lors, un ingénieur de Facebook s'est interrogé[35] sur deux aspects :

- quel intérêt aurait un acteur à jouer le rôle d'émetteur ?
- quels risques ce rôle ferait peser sur le respect de la vie privée des utilisateurs du site de cet acteur émetteur ? Posséder un jeton émis par un acteur divulgue à minima que l'internaute utilise un service de cet acteur.

Pour rester attractifs face aux annonceurs soucieux de ne pas afficher inutilement leurs publicités à des robots, les éditeurs pourraient devenir dépendants des jetons d'authenticité. En pratique, cela renforcerait le rôle des sites émetteurs et pourrait créer une forme de concentration comme ce qui a pu être observé avec les modules d'authentification type oAuth[36]. Cela ajouterait potentiellement une dépendance en raison du risque de baisse de valeur de la publicité pour les sites qui n'intégreraient pas ces mécanismes. Pour ces raisons, le marché de l'authenticité en ligne nécessitera une vigilance particulière afin de prévenir d'éventuels abus ou dérives.

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

En février 2022, Google a annoncé vouloir étendre le projet de la *Privacy Sandbox* à l'univers mobile. L'objectif était similaire au projet web : permettre de fournir aux applications des revenus via la publicité tout en limitant le *tracking* et notamment l'usage de l'*Advertising ID*, un identifiant publicitaire unique[37] associé au terminal de l'utilisateur.

Les projets web et Android, sous une direction unique chez Google, partagent des outils tels que l'*API Topics* et l'*API Protected Audience*. Bien qu'adaptés au fonctionnement des applications, ces outils garderont une identité commune. Contrairement à sa version web, le projet *Privacy*

[33] *Private State Tokens Issuer Registration* sur Github : https://github.com/GoogleChrome/private-tokens/blob/main/PST-Registration.md#private-state-tokens-issuer-registration

[34] Chromium PST Demo, hCaptcha, Polyset, CaptchaFox et AUTHFY ont complété l'enregistrement au 23 février 2024, sur Github : https://github.com/GoogleChrome/private-tokens/issues?q=is%3Aissue

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

[36] OAuth est un protocole libre qui permet d'autoriser un site web, un logiciel ou une application (dite « consommateur ») à utiliser l'API sécurisée d'un autre site web (dit « fournisseur ») pour le compte d'un utilisateur. OAuth n'est pas un protocole d'authentification, mais de « délégation d'autorisation ». Des fournisseurs connus étant Facebook, Google ou encore Apple : https://fr.wikipedia.org/wiki/OAuth

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

Sandbox pour Android n'est pas partagé sur une plateforme collaborative tierce, comme Github. La participation aux outils se fait directement sur le site web d'Android[38], propriété de Google[39], via des formulaires. Cette procédure rend la construction des outils de la *Privacy Sandbox* sous Android moins facile à suivre que son homologue web.

Tout comme le retrait des *cookies* tiers sur son équivalent web, la *Privacy Sandbox* sous Android s'accompagne d'une innovation en termes de protection de la vie privée : *SDK Runtime*[40]. Cet outil vise à reconfigurer l'intégration des kits de développement logiciel *(Software Development Kit* - SDK[41]) au sein des applications Android.

Ces kits sont fréquemment utilisés pour ajouter des fonctionnalités à une application (par exemple, ajout d'un SDK publicitaire à des fins de monétisation). Sous Android, un SDK embarqué par une application mobilise le même espace de mémoire vive que cette dernière et peut donc accéder aux informations qu'elle exploite pour son fonctionnement. Ainsi, lorsqu'une application a accès à la localisation de l'utilisateur (pour un GPS par exemple), et que cette application embarque un ou plusieurs SDKs, ces derniers peuvent accéder aux informations de localisation de l'utilisateur, sans nécessiter l'aval du développeur de l'application.

SDK Runtime introduit la technologie des *« Runtime-Enabled SDK »* (RE-SDK) qui permet la séparation entre la mémoire vive utilisée d'une application de celles de ses RE-SDKs associés. Les RE-SDKs communiquent alors avec l'application par le biais d'APIs spécifiques, sans avoir accès aux informations détenues par l'application.

Ce nouveau fonctionnement des RE-SDKs permet également une distribution différente des applications. Les développeurs d'applications n'ont pas besoin d'embarquer les kits avec le code de leur propre application lors de la publication sur le *Play Store* mais peuvent simplement lister les SDKs que leur application utilise avec les numéros de version associés. Ce système a l'avantage de potentiellement mutualiser un SDK pour plusieurs applications installées, ce qui réduit l'espace mémoire occupé sur le terminal de l'utilisateur. Cela permet également une distribution plus aisée des mises-à-jour des SDKs, sans intervention des développeurs d'application[42].

Néanmoins, contrairement à la disparition des *cookies* tiers sur le navigateur Chrome, l'utilisation des RE-SDKs s'annonce facultative et pour le moment s'adresse en priorité aux SDKs publicitaires.

---

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

[39] Page « *Brand Guidelines* » de Android : https://developer.android.com/distribute/marketing-tools/brand-guidelines

[40] SDK Runtime, Android Developer : https://developer.android.com/design-for-safety/privacy-sandbox/sdk-runtime

[41] Définition d'un SDK, Glossaire Mozilla : https://developer.mozilla.org/en-US/docs/Glossary/SDK

[42] Actuellement, lorsqu'un SDK est mis-à-jour, les développeurs d'applications qui utilisent ce SDK doivent prendre en compte cette mise-à-jour au sein de leur application. Avec le nouveau système, une mise-à-jour mineure ne nécessitera pas d'intervention d'un développeur d'application.

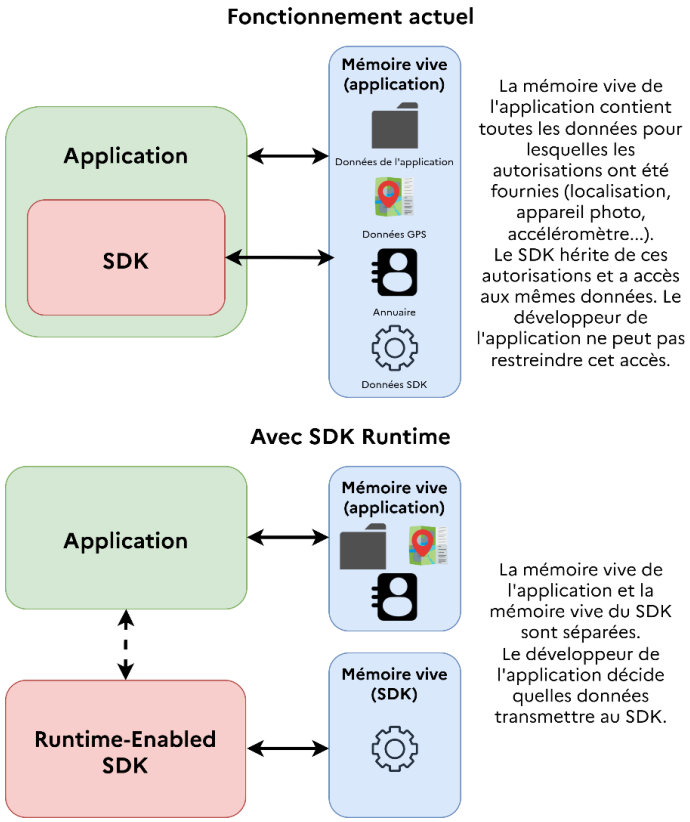

Figure 5 : Illustration du principe de séparation de la mémoire grâce à SDK Runtime

PRIVACY SANDBOX : DYNAMIQUE DU PROJET

C'est en réponse à l'enjeu grandissant de protection de la vie privée et à différents scandales[43] que Mozilla et Apple ont ouvert la voie en limitant l'accès aux données de leurs utilisateurs, au détriment des revenus de l'industrie[44] publicitaire.

En 2019, Google leur emboîte le pas en initiant 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 progressivement à 2024 puis 2025. Chaque outil du projet *Privacy Sandbox* est décrit publiquement, et la liste de ces outils est disponible sur le site de présentation du projet[45]. Tous ces outils appartiennent à différentes catégories telles que « Affichage de publicités et de contenus pertinents », « Mesures des annonces numériques » ou encore « Lutte contre le spam et la fraude sur le web[46] ».

[43] *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/

[44] Op. cit. 10. 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).

[45] Op. cit. 11.

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

Au cours du premier semestre 2020, un florilège d'outils de la *Privacy Sandbox* passent rapidement en phase d'incubation, certains atteignant la phase d'expérimentation dès 2021. Dès ces premières expérimentations, des doutes sont exprimés[47] sur les conséquences des technologies promues par Google et leurs effets sur la libre concurrence en matière de publicité en ligne. En juin 2021, une annonce de procédure négociée est publiée par l'autorité britannique de la concurrence, la *Competition and Markets Authority* (CMA[48]), craignant notamment que la *Privacy Sandbox* ne renforce la position de Google dans l'écosystème de la publicité[49]. Google prend de premiers engagements et promet plus de transparence autour de l'initiative. Ainsi, un calendrier prévisionnel officiel est finalement publié pour le développement et les tests des différents outils[50]. En novembre 2021, après une consultation par la CMA des acteurs du secteur, Google renforce ses engagements, sans répondre complètement aux inquiétudes émises par l'autorité britannique[51].

En parallèle, certains acteurs tels que Criteo, spécialiste français de la publicité ciblée, ont participé au développement des outils de la *Privacy Sandbox* et aux phases de tests qui l'ont accompagné, publiant des articles sur les résultats de leurs expérimentations[52]. Pour le moment, les résultats obtenus annoncent une baisse de revenus en comparaison des technologies associées aux *cookies*.

En dépit d'une volonté de développer des standards, la *Privacy Sandbox* ne fait pas l'unanimité auprès des navigateurs concurrents, jusqu'ici réticents à adopter ses outils. Par exemple, la proposition d'intérêts basés sur l'historique, *Topics,* a reçu des avis négatifs de la part de Mozilla[53] et Webkit[54] (le moteur de Safari). Le *Technical Architecture Group* ou TAG[55], un groupe du W3C, a soulevé quant à lui des inquiétudes[56] en termes de fragmentation du web et d'évolution du rôle du navigateur.

Toutefois, certains des outils de protection de la vie privée font partie des rares à avoir eu des retours positifs : c'est par exemple le cas de

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

[48] Competition and Markets 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

[49] Voir par exemple le point 5.3 du rapport de la CMA (Op. cit., 48).

[50] Op. cit. 13.

[51] 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

[52] Medium, articles de Criteo au sujet de la *Privacy Sandbox :* https://medium.com/criteo-engineering/tagged/privacy-sandbox

[53] Position de Mozilla sur Topics, GitHub : https://github.com/mozilla/standards-positions/issues/622

[54] Position de Webkit sur Topics, GitHub : https://github.com/WebKit/standards-positions/issues/111

[55] Le TAG est un groupe de travail du W3C, ses missions sont décrites plus en détail ici : https://www.w3.org/2001/tag/. C'est ce groupe qui s'occupe d'étudier les solutions de ciblage de la *Privacy Sandbox*

[56] Voir ces messages : https://github.com/w3ctag/design-reviews/issues/726#issuecomment-1379908459 et https://github.com/w3ctag/design-reviews/issues/726#issuecomment-1612522047

CHIPS[57] (*Cookies Having Independent Partitioned State*), qui a reçu un avis positif de Mozilla[58]. Webkit[59] est prêt à encourager l'initiative si de légères modifications sont apportées.

Google assure que ses différentes API limiteront le pistage abusif des internautes et leur ré-identification (c'est-à-dire la reconnaissance d'un utilisateur sans son consentement). L'utilisation des API de la *Privacy Sandbox* est tout de même soumise à un enregistrement préalable auprès de Google. Gratuit et rapide, cet enregistrement engage les annonceurs ou éditeurs à ne pas abuser des outils de la *Privacy Sandbox* à des fins de ré-identification entre différents sites. Cet engagement prend la forme d'une attestation publiée sur leur site web et téléchargée automatiquement par les navigateurs. Google a annoncé vouloir rendre publique la liste des utilisateurs des API de la *Privacy Sandbox*. En attente, cette parution offrirait un surcroît de transparence sur le fonctionnement de l'écosystème, tout en étant un *nudge* potentiel pour le marché si Google arrive à convaincre éditeurs et utilisateurs de la pertinence de ses nouveaux outils.

CONSÉQUENCES DE LA *PRIVACY SANDBOX* SUR LA VALEUR DE LA PUBLICITÉ EN LIGNE

La capacité de la solution *Privacy Sandbox* à monétiser[60] aussi bien que les technologies de *cookies* n'est pas démontrée à ce jour. 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 ou que celle-ci soit mieux captée par les acteurs les plus gros.

Les acteurs œuvrant notamment dans des environnements exigeant une authentification de l'utilisateur seront moins affectés par ce changement technologique qui n'entrave en rien leur capacité à collecter des données étayées via des *cookies first-party*. Ces acteurs ne seront aucunement désavantagés et connaîtront au contraire une concurrence moindre. En pratique, les acteurs disposant de nombreuses données sur leurs utilisateurs sont souvent des réseaux sociaux : il est nécessaire d'être connecté pour y accéder, et leur utilisation repose sur la publication ou la consommation de contenus qui dévoilent des centres d'intérêt, des pratiques de consommation, etc. À l'inverse, les éditeurs qui publient leur contenu sans nécessiter d'authentification (notamment les sites de presse) risquent de voir la valeur de leurs encarts publicitaires baisser.

[57] CHIPS permet de partitionner des cookies tiers : c'est-à-dire que sur un site donné, un acteur tiers peut toujours déposer un cookie tiers. Cependant, ce cookie ne sera pas lisible par l'acteur tiers lorsque l'internaute naviguera sur un autre site. Ainsi, les acteurs peuvent stocker de l'information à propos d'un utilisateur sur un site donné, mais ils ne peuvent pas le suivre pour le reste de sa navigation.

[58] Position de Mozilla sur CHIPS, GitHub : https://github.com/mozilla/standards-positions/issues/678

[59] Position de Webkit sur CHIPS, GitHub : https://github.com/WebKit/standards-positions/issues/50

[60] 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

Depuis l'annonce de la disparition des *cookies* tiers qui promet un changement majeur pour le marché de la publicité en ligne, certains acteurs ont travaillé sur des projets alternatifs. Certaines solutions peuvent être respectueuses de la vie privée, tandis que d'autres permettent de continuer à suivre de près les parcours des utilisateurs web sans recours aux *cookies* tiers. Ces dernières peuvent être basées sur l'utilisation d'un identifiant unique (comme une adresse mail ou un numéro de téléphone), ou sur des techniques de *fingerprinting*… Ces outils pourraient venir compléter ou concurrencer les outils de la *Privacy Sandbox*.

Une baisse des revenus publicitaires pourrait pousser certains sites à se tourner vers des solutions identifiantes, ce qui irait à l'encontre de l'objectif de minimisation de la collecte de données personnelles.

Enfin, face à la disparition imminente des *cookies* tiers, des témoignages affluent faisant état de la difficulté croissante à adopter les outils de la *Privacy Sandbox*. Le fait d'avoir repoussé à plusieurs reprises le retrait des *cookies* tiers ne semble pas pour autant avoir donné des périodes de tests fructueuses et favoriser l'adoption par le plus grand nombre. Des fenêtres d'essais avec trop peu d'utilisateurs et la concurrence active des *cookies* tiers toujours existants limitent les études en conditions réelles.

Cette année 2024 devrait néanmoins révéler les évolutions du marché de la publicité en ligne, entre adoption plus large de la *Privacy Sandbox* et solutions alternatives.

La collection « Éclairage sur… » du PEReN propose, dans un format didactique, des éléments d'analyse techniques sur des thèmes liés à la régulation des plateformes numériques. Retrouvez l'ensemble des numéros parus à l'adresse www.peren.gouv.fr/publications/

Dépôt légal : Octobre 2022
ISSN (en ligne) : 2824-8201

Service à compétence nationale placé sous la tutelle des Ministres de l'économie, du numérique et de la Culture, 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 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 - contact.peren@finances.gouv.fr