

# Les données que récolte Google – Ch.4

*Voici déjà la traduction du quatrième chapitre de [Google Data Collection](#), l'étude élaborée par l'équipe du professeur Douglas C. Schmidt, spécialiste des systèmes logiciels, chercheur et enseignant à l'[Université Vanderbilt](#). Si vous les avez manqués, retrouvez les [chapitres précédents déjà publiés](#). Il s'agit cette fois d'explorer les stratégies des régies publicitaires qui opèrent en arrière-plan : des opérations fort discrètes mais terriblement efficaces...*

*Traduction Framalang : Côme, goofy, Khrys,Obny, Penguin, Piup, serici.*

## IV. Collecte de données par les outils des annonceurs et des diffuseurs

29. Une source majeure de collecte des données d'activité des utilisateurs provient des outils destinés au annonceurs et aux éditeurs tels que Google Analytics, DoubleClick, AdSense, AdWords et AdMob. Ces outils ont une portée énorme ; par exemple, plus d'un million d'applications mobiles utilisent AdMob<sup>1</sup>, plus d'un million d'annonceurs utilisent AdWords<sup>2</sup>, plus de 15 millions de sites internet utilisent AdSense<sup>3</sup> et plus de 30 millions de sites utilisent Google Analytics<sup>4</sup>.

30. Au moment de la rédaction du présent rapport, Google a rebaptisé AdWords « *Google Ads* » et DoubleClick « *Google Ad Manager* », mais aucune modification n'a été apportée aux fonctionnalités principales des produits, y compris la collecte d'informations par ces produits<sup>5</sup>. Par conséquent, pour

les besoins du présent rapport, les premiers noms ont été conservés afin d'éviter toute confusion avec des noms de domaine connexes (tels que doubleclick.net).

31. Voici deux principaux groupes d'utilisateurs des outils de Google axés sur l'édition – et les annonces publicitaires :

- Les éditeurs de sites web et d'applications, qui sont des organisations qui possèdent des sites web et créent des applications mobiles. Ces entités utilisent les outils de Google pour (1) gagner de l'argent en permettant l'affichage d'annonces aux visiteurs sur leurs sites web ou applications, et (2) mieux suivre et comprendre qui visite leurs sites et utilise leurs applications. Les outils de Google placent des cookies et exécutent des scripts dans les navigateurs des visiteurs du site web pour aider à déterminer l'identité d'un utilisateur et suivre son intérêt pour le contenu et son comportement en ligne. Les bibliothèques d'applications mobiles de Google suivent l'utilisation des applications sur les téléphones mobiles.
- Les annonceurs, qui sont des organisations qui paient pour que des bannières, des vidéos ou d'autres publicités soient diffusées aux utilisateurs lorsqu'ils naviguent sur Internet ou utilisent des applications. Ces entités utilisent les outils de Google pour cibler des profils spécifiques de personnes pour que les publicités augmentent le retour sur leurs investissements marketing (les publicités mieux ciblées génèrent généralement des taux de clics et de conversion plus élevés). De tels outils permettent également aux annonceurs d'analyser leurs audiences et de mesurer l'efficacité de leur publicité numérique en regardant sur quelles annonces les utilisateurs cliquent et à quelle fréquence, et en donnant un aperçu du profil des personnes qui ont cliqué sur les annonces.

32. Ensemble, ces outils recueillent des informations sur les

activités des utilisateurs sur les sites web et dans les applications, comme le contenu visité et les annonces cliquées. Ils travaillent en arrière-plan – en général imperceptibles par des utilisateurs. La figure 7 montre certains de ces outils clés, avec des flèches indiquant les données recueillies auprès des utilisateurs et les publicités qui leur sont diffusées.

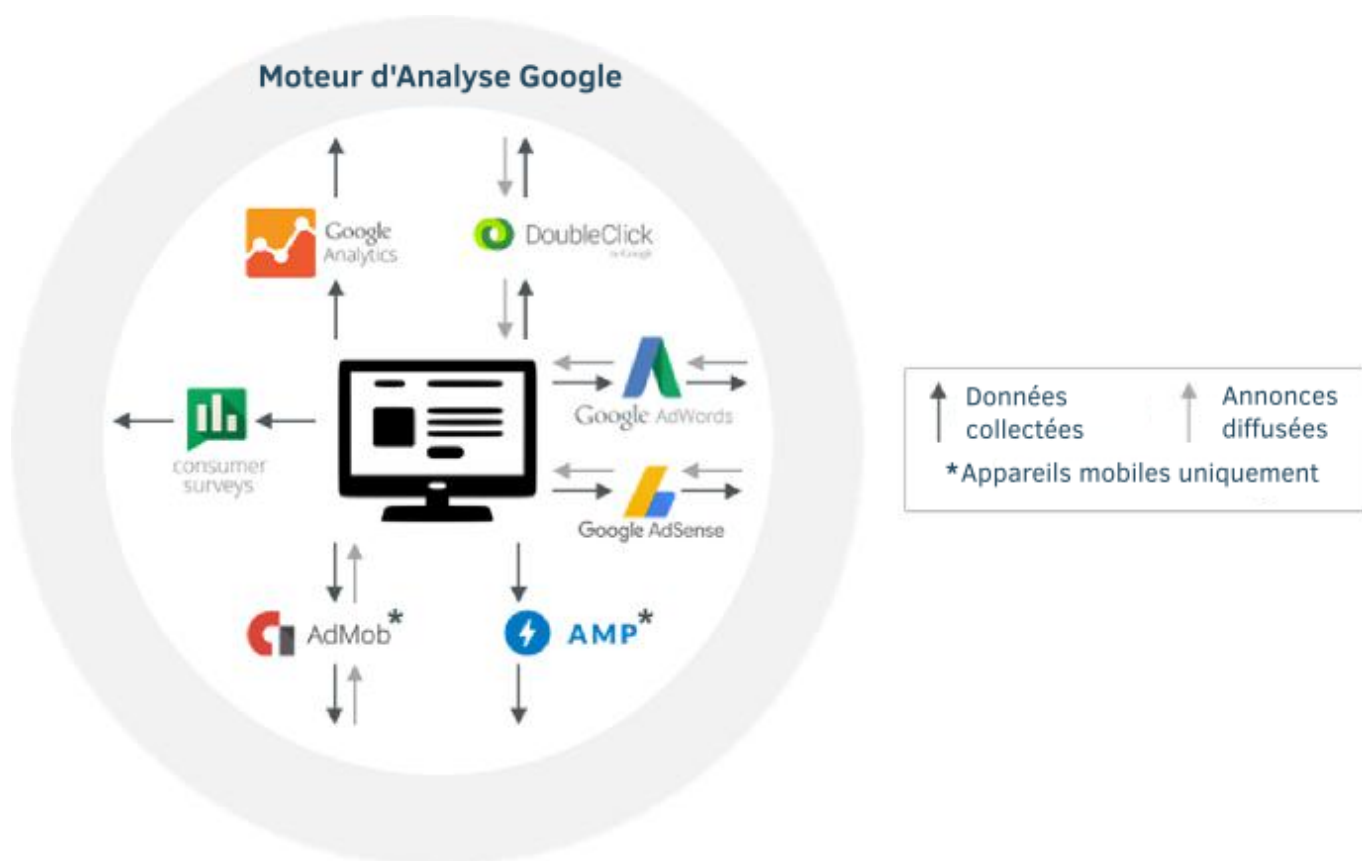


Figure 7 : Produits Google destinés aux éditeurs et annonceurs<sup>6</sup>

33. Les informations recueillies par ces outils comprennent un identifiant non personnel que Google peut utiliser pour envoyer des publicités ciblées sans identifier les informations personnelles de la personne concernée. Ces identificateurs peuvent être spécifiques à l'appareil ou à la session, ainsi que permanents ou semi-permanents. Le tableau 1 liste un ensemble de ces identificateurs. Afin d'offrir aux utilisateurs un plus grand anonymat lors de la collecte d'informations pour le ciblage publicitaire, Google s'est récemment tourné vers l'utilisation d'identifiants uniques

semi-permanents (par exemple, les GAID)<sup>7</sup>. Des sections ultérieures décrivent en détail la façon dont ces outils recueillent les données des utilisateurs et l'utilisation de ces identificateurs au cours du processus de collecte des données.

**Tableau 1: Identificateurs transmis à Google**

<b>Identificateur</b>	<b>Type</b>	<b>Description</b>
GAID/IDFA	Semi-permanent	Chaine de caractères alphanumériques pour appareils Android et iOS, pour permettre les publicités ciblées sur mobile. Réinitialisable par l'utilisateur.
ID client	Semi-permanent	ID créé la première fois qu'un cookie est stocké sur le navigateur. Utilisé pour relier les sessions de navigations. Réinitialisé lorsque les cookies du navigateur sont effacés.
Adresse IP	Semi-permanent	Une unique suite de nombre qui identifie le réseau par lequel un appareil accède à internet.

ID appareil Android	Semi-permanent	Nombre généré aléatoirement au premier démarrage d'un appareil. Utilisé pour identifier l'appareil. En retrait progressif pour la publicité. Réinitialisé lors d'une remise à zéro de l'appareil.
Google Services Framework (GSF)	Semi-permanent	Nombre assigné aléatoirement lorsqu'un utilisateur s'enregistre pour la première fois dans les services Google sur un appareil. Utilisé pour identifier un appareil unique. Réinitialisé lors d'une remise à zéro de l'appareil.
IEMI / MEID	Permanent	Identificateur utilisé dans les standards de communication mobile. Unique pour chaque téléphone portable.
Adresse MAC	Permanent	Identificateur unique de 12 caractères pour un élément matériel (ex. : routeur).
Numéro de série	Permanent	Chaine de caractères alphanumériques utilisée pour identifier un appareil.

## A. Google Analytics et DoubleClick

34. DoubleClick et Google Analytics (GA) sont les produits

phares de Google en matière de suivi du comportement des utilisateurs et d'analyse du trafic des pages Web sur les périphériques de bureau et mobiles. GA est utilisé par environ 75 % des 100 000 sites Web les plus visités<sup>8</sup>. Les cookies DoubleClick sont associés à plus de 1,6 million de sites Web<sup>9</sup>.

35. GA utilise de petits segments de code de traçage (appelés « balises de page ») intégrés dans le code HTML d'un site Web<sup>10</sup>. Après le chargement d'une page Web à la demande d'un utilisateur, le code GA appelle un fichier *analytics.js* qui se trouve sur les serveurs de Google. Ce programme transfère un instantané « par défaut » des données de l'utilisateur à ce moment, qui comprend l'adresse de la page web visitée, le titre de la page, les informations du navigateur, l'emplacement actuel (déduit de l'adresse IP), et les paramètres de langue de l'utilisateur. Les scripts de GA utilisent des cookies pour suivre le comportement des utilisateurs.

36. Le script de GA, la première fois qu'il est exécuté, génère et stocke un cookie spécifique au navigateur sur l'ordinateur de l'utilisateur. Ce cookie a un identificateur de client unique (voir le tableau 1 pour plus de détails)<sup>11</sup> Google utilise l'identificateur unique pour lier les cookies précédemment stockés, qui capturent l'activité d'un utilisateur sur un domaine particulier tant que le cookie n'expire pas ou que l'utilisateur n'efface pas les cookies mis en cache dans son navigateur<sup>12</sup>

37. Alors qu'un cookie GA est spécifique au domaine particulier du site Web que l'utilisateur visite (appelé « cookie de première partie »), un cookie DoubleClick est généralement associé à un domaine tiers commun (tel que doubleclick.net). Google utilise de tels cookies pour suivre l'interaction de l'utilisateur sur plusieurs sites web tiers<sup>13</sup> Lorsqu'un utilisateur interagit avec une publicité sur un site

web, les outils de suivi de conversion de DoubleClick (par exemple, Floodlight) placent des cookies sur l'ordinateur de l'utilisateur et génèrent un identifiant client unique<sup>14</sup> Par la suite, si l'utilisateur visite le site web annoncé, le serveur DoubleClick accède aux informations stockées dans le cookie, enregistrant ainsi la visite comme une conversion valide.

## **B. AdSense, AdWords et AdMob**

38. AdSense et AdWords sont des outils de Google qui diffusent des annonces sur les sites Web et dans les résultats de recherche Google, respectivement. Plus de 15 millions de sites Web ont installé AdSense pour afficher des annonces sponsorisées<sup>15</sup> De même, plus de 2 millions de sites web et applications, qui constituent le réseau Google Display Network (GDN) et touchent plus de 90 % des internautes<sup>16</sup> affichent des annonces AdWords.

39. AdSense collecte des informations indiquant si une annonce a été affichée ou non sur la page web de l'éditeur. Il recueille également la façon dont l'utilisateur a interagi avec l'annonce, par exemple en cliquant sur l'annonce ou en suivant le mouvement du curseur sur l'annonce<sup>17</sup>. AdWords permet aux annonceurs de diffuser des annonces de recherche sur Google Search, d'afficher des annonces sur les pages des éditeurs et de superposer des annonces sur des vidéos YouTube. Pour suivre les taux de clics et de conversion des utilisateurs, les publicités AdWords placent un cookie sur les navigateurs des utilisateurs pour identifier l'utilisateur s'il visite par la suite le site web de l'annonceur ou s'il effectue un achat<sup>18</sup>.

40. Bien qu'AdSense et AdWords recueillent également des données sur les appareils mobiles, leur capacité d'obtenir des renseignements sur les utilisateurs des appareils mobiles est limitée puisque les applications mobiles ne partagent pas de

cookies entre elles, une technique d'isolement appelée « bac à sable »<sup>19</sup> qui rend difficile pour les annonceurs de suivre le comportement des utilisateurs entre différentes applications mobiles.

41 Pour résoudre ce problème, Google et d'autres entreprises utilisent des « bibliothèques d'annonces » mobiles (comme AdMob) qui sont intégrées dans les applications par leurs développeurs pour diffuser des annonces dans les applications mobiles. Ces bibliothèques compilent et s'exécutent avec les applications et envoient à Google des données spécifiques à l'application à laquelle elles sont intégrées, y compris les emplacements GPS, la marque de l'appareil et le modèle de l'appareil lorsque les applications ont les autorisations appropriées. Comme on peut le voir dans les analyses de trafic de données (Figure 8), et comme on peut trouver confirmation sur les propres pages web des développeurs de Google<sup>20</sup>, de telles bibliothèques peuvent également envoyer des données personnelles de l'utilisateur, telles que l'âge et le genre, tout cela va vers Google à chaque fois que les développeurs d'applications envoient explicitement leurs valeurs numériques vers la bibliothèque.

```
platform=LGE
submodel=LGUS610
rm=1
android_app_muted=false
request_id=aeca1769-9f28-42f0-98e6-fdb92ff796a0
am=0
cnt=1
ma=0
disable_ml=false
js=afma-sdk-a-v12673021.11910000.1
session_id=12059440741457373925
muv=7
cust_gender=2
```

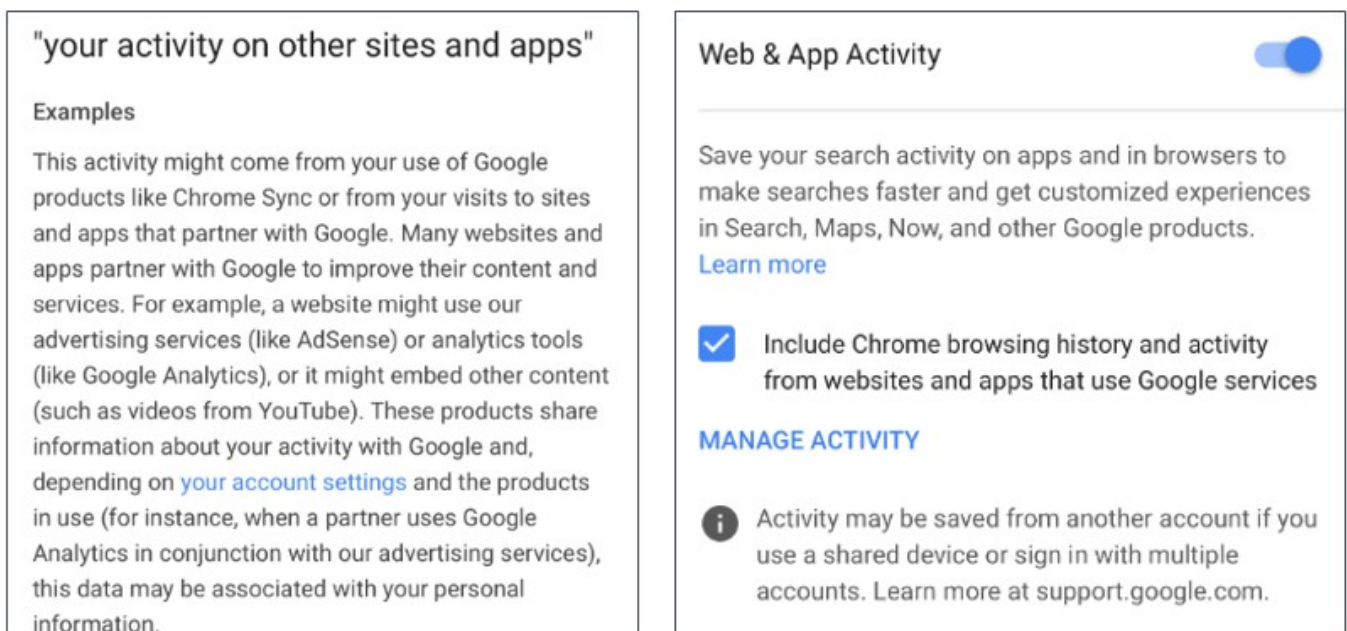
Genre de l'utilisateur →

Figure 8 : Aperçu des informations renvoyées à Google lorsqu'une application est lancée



## C. Association de données recueillies passivement et d'informations à caractère personnel

42. Comme nous l'avons vu plus haut, Google recueille des données par l'intermédiaire de produits pour éditeurs et annonceurs, et associe ces données à une variété d'identificateurs semi-permanents et anonymes. Google a toutefois la possibilité d'associer ces identifiants aux informations personnelles d'un utilisateur. C'est ce qu'insinuent les déclarations faites dans la politique de confidentialité de Google, dont des extraits sont présentés à la figure 9. La zone de texte à gauche indique clairement que Google peut associer des données provenant de services publicitaires et d'outils d'analyse aux informations personnelles d'un utilisateur, en fonction des paramètres du compte de l'utilisateur. Cette disposition est activée par défaut, comme indiqué dans la zone de texte à droite.



The image shows two side-by-side panels. The left panel is a text box with a black border containing the following text:

"your activity on other sites and apps"

Examples

This activity might come from your use of Google products like Chrome Sync or from your visits to sites and apps that partner with Google. Many websites and apps partner with Google to improve their content and services. For example, a website might use our advertising services (like AdSense) or analytics tools (like Google Analytics), or it might embed other content (such as videos from YouTube). These products share information about your activity with Google and, depending on [your account settings](#) and the products in use (for instance, when a partner uses Google Analytics in conjunction with our advertising services), this data may be associated with your personal information.

The right panel is a screenshot of the 'Web & App Activity' settings page. At the top, the title 'Web & App Activity' is followed by a blue toggle switch that is turned on. Below the title, there is a horizontal line. The main text reads: 'Save your search activity on apps and in browsers to make searches faster and get customized experiences in Search, Maps, Now, and other Google products.' Below this is a blue link 'Learn more'. A blue checkmark icon is followed by the text: 'Include Chrome browsing history and activity from websites and apps that use Google services'. Below this is a blue link 'MANAGE ACTIVITY'. At the bottom, there is an information icon (i) followed by the text: 'Activity may be saved from another account if you use a shared device or sign in with multiple accounts. Learn more at support.google.com.'

Figure 9 : Page de confidentialité de Google pour la collecte de sites web tiers et l'association avec des informations personnelles<sup>2122</sup>.

43. De plus, une analyse du trafic de données échangé avec les serveurs de Google (résumée ci-dessous) a permis d'identifier deux exemples clés (l'un sur Android et l'autre sur Chrome) qui montrent la capacité de Google à corréler les données recueillies de façon anonyme avec les renseignements personnels des utilisateurs.

### **1) L'identificateur de publicité mobile peut être désanonymé grâce aux données envoyées à Google par Android.**

44. Les analyses du trafic de données communiqué entre un téléphone Android et les domaines de serveur Google suggèrent un moyen possible par lequel des identifiants anonymes (GAID dans ce cas) peuvent être associés au compte Google d'un utilisateur. La figure 10 décrit ce processus en une série de trois étapes clés.

45. Dans l'étape 1, une donnée de *check-in* est envoyée à l'URL `android.clients.google.com/checkin`. Cette communication particulière fournit une synchronisation de données Android aux serveurs Google et contient des informations du journal Android (par exemple, du journal de récupération), des messages du noyau, des crash dumps, et d'autres identifiants liés au périphérique. Un instantané d'une demande d'enregistrement partiellement décodée envoyée au serveur de Google à partir d'Android est montré en figure 10.

### 1 Checkin upload

MAC address, IMEI / MEID, Gmail account →

```
locale: "en-US"
loggingId: -5117406526731365***
macAddr: "2c598a23****"
meid: "35956507040****"
accountCookie: "[jane.gcollpro@gmail.com]"
accountCookie: "ya29.GokBuQWfVoYx9I9S/S2Hxksk-K0CZMUvVw1_hZO"
timeZone: "America/New_York"
securityToken: 2869396553181660***
version: 3
otaCert: "EGSUn3Y+V7PXdlJYmV9laSOFcic="
Serial number → serialNumber: "LGUS610b12c****"
deviceConfiguration {
  touchScreen: 3
  keyboard: 1
  navigation: 1
  screenLayout: 2
  hasHardKeyboard: false
  hasFiveWayNavigation: false
  screenDensity: 320
```

### 2 Google's checkin response

```
1: 1
2: {
  1: com.google.android.gms.checkin.NOOP
}
3: 1527172739355
5: {
  1: cw:companion_logging_buffer_size
  2: {
  }
}
}
}
Google Service Framework (decimal number) → 7: 412694972786808**** → Hex number: 3945DEE97778****
8: 6301475365587305549
9: 1
12: ABFEt1XH4UAKVDgXQNwDo_8AKaZTux9pGRAU1g1RdW2vdLHYpp)
```

### 3 Android bulk details

Google Services Framework (hex number) →

```
:authority android.clients.google.com
content-length 1205
user-agent Android-Finsky/10.1.08-all%20%5B0%5D%20%5BPI
x-dfe-content-filters
x-dfe-device-id 3945dee97778****
if-none-match -1.6E+09
x-dfe-device-config-token CisaKQoTNDEyNjkOOTcyNzg2ODA4MzlyMRISChAxf
authorization GoogleLogin auth=IwZo4lcwvvnwk8QKixU7lu-r7wh
accept-language en-US
x-dfe-cookie CAEQARgAlgJVuygBMjxDaXNhS1FvVESERXIOamsw
x-limit-ad-tracking-enabled FALSE
x-dfe-device-checkin-consistency-token ABFEt1U16JQ5xikF9j8u3RyZGjAlS947Qmb3sYJ098-V
x-dfe-network-type 3
x-dfe-mccmnc 310260
x-dfe-client-id am-android-uscellular-us
x-dfe-userlanguages en_US
x-dfe-phenotype H4sIAAAAAAAAAABXPUW6CMAAAD0x3n7uHCaggJv
x-dfe-encoded-targets CAESxwJX3JSBBg6mCtECDzMC2AMBARSSB4AcPAtI
x-dfe-request-params timeoutMs=30000
GAID → x-ad-id 83d479a5-5eb7-44f1-a2ee-14e91094****
content-type application/x-protobuf
accept-encoding gzip, deflate
```

Figure 10 : Les identifiants d'appareil sont envoyés avec les informations de compte dans les requêtes de vérification Android.

46. Comme l'indiquent les zones pointées, Android envoie à Google, au cours du processus d'enregistrement, une variété d'identifiants permanents importants liés à l'appareil, y compris l'adresse MAC de l'appareil, l'IMEI /MEID et le numéro de série du dispositif. En outre, ces demandes contiennent également l'identifiant Gmail de l'utilisateur Android, ce qui permet à Google de relier les informations personnelles d'un utilisateur aux identifiants permanents des appareils Android.

47. À l'étape 2, le serveur de Google répond à la demande d'enregistrement. Ce message contient un identifiant de cadre de services Google (GSF ID)<sup>23</sup> qui est similaire à l'« Android ID »<sup>24</sup> (voir le tableau 1 pour les descriptions).

48. L'étape 3 implique un autre cas de communication où le même identifiant GSF (de l'étape 2) est envoyé à Google en même temps que le GAID. La figure 10 montre l'une de ces transmissions de données à `android.clients.google.com/fdfe/bulkDetails?au=1`.

49. Grâce aux trois échanges de données susmentionnés, Google reçoit les informations nécessaires pour connecter un GAID avec des identifiants d'appareil permanents ainsi que les identifiants de compte Google des utilisateurs.

50. Ces échanges de données interceptés avec les serveurs de Google à partir d'un téléphone Android montrent comment Google peut connecter les informations anonymisées collectées sur un appareil mobile Android via les outils DoubleClick, Analytics ou AdMob avec l'identité personnelle de l'utilisateur. Au cours de la collecte de données sur 24 heures à partir d'un téléphone Android sans mouvement ni activité, deux cas de communications d'enregistrement avec des serveurs Google ont été observés. Une analyse supplémentaire est toutefois nécessaire pour déterminer si un tel échange d'informations a lieu avec une certaine périodicité ou s'il est déclenché par des activités spécifiques sur les téléphones.

## **2) L'ID du cookie DoubleClick est relié aux informations personnelles de l'utilisateur sur le compte Google.**

51. La section précédente expliquait comment Google peut désanonymiser l'identité de l'utilisateur via les données passives et anonymisées qu'il collecte à partir d'un appareil mobile Android. Cette section montre comment une telle désanonymisation peut également se produire sur un ordinateur de bureau/ordinateur portable.

52. Les données anonymisées sur les ordinateurs de bureau et portables sont collectées par l'intermédiaire d'identifiants basés sur des cookies (par ex. Cookie ID), qui sont

typiquement générés par les produits de publicité et d'édition de Google (par ex. DoubleClick) et stockés sur le disque dur local de l'utilisateur. L'expérience présentée ci-dessous a permis d'évaluer si Google peut établir un lien entre ces identificateurs (et donc les renseignements qui y sont associés) et les informations personnelles d'un utilisateur. Cette expérience comportait les étapes ordonnées suivantes :

1. Ouverture d'une nouvelle session de navigation (Chrome ou autre) (pas de cookies enregistrés, par exemple navigation privée ou incognito) ;
2. Visite d'un site Web tiers qui utilisait le réseau publicitaire DoubleClick de Google ;
3. Visite du site Web d'un service Google largement utilisé (Gmail dans ce cas) ;
4. Connexion à Gmail.

53. Au terme des étapes 1 et 2, dans le cadre du processus de chargement des pages, le serveur DoubleClick a reçu une demande lorsque l'utilisateur a visité pour la première fois le site Web tiers. Cette demande faisait partie d'une série de requêtes comprenant le processus d'initialisation DoubleClick lancé par le site Web de l'éditeur, qui a conduit le navigateur Chrome à installer un cookie pour le domaine DoubleClick. Ce cookie est resté sur l'ordinateur de l'utilisateur jusqu'à son expiration ou jusqu'à ce que l'utilisateur efface manuellement les cookies via les paramètres du navigateur.

54. Ensuite, à l'étape 3, lorsque l'utilisateur visite Gmail, il est invité à se connecter avec ses identifiants Google. Google gère l'identité à l'aide d'une architecture single sign on (SSO) [NdT : authentification unique], dans laquelle les identifiants sont fournis à un service de compte (ici `accounts.google.com`) en échange d'un « jeton d'authentification », qui peut ensuite être présenté à d'autres services Google pour identifier les utilisateurs. À l'étape 4, lorsqu'un utilisateur accède à son compte Gmail, il

se connecte effectivement à son compte Google, qui fournit alors à Gmail un jeton d'autorisation pour vérifier l'identité de l'utilisateur.<sup>25</sup> Ce processus est décrit à la figure 24 de la section IX.E de l'annexe.

55. Dans la dernière étape de ce processus de connexion, une requête est envoyée au domaine DoubleClick. Cette requête contient à la fois le jeton d'authentification fourni par Google et le cookie de suivi défini lorsque l'utilisateur a visité le site web tiers à l'étape 2 (cette communication est indiquée à la figure 11). Cela permet à Google de relier les informations d'identification Google de l'utilisateur à un cookie DoubleClick. Par conséquent, si les utilisateurs n'effacent pas régulièrement les cookies de leur navigateur, leurs informations de navigation sur les pages Web de tiers qui utilisent les services DoubleClick pourraient être associées à leurs informations personnelles sur Google Account.



Figure 11 : La requête à DoubleClick.net inclut le jeton d'authentification Google et les cookies passés.

56. Il est donc établi à présent que Google recueille une grande variété de données sur les utilisateurs par l'intermédiaire de ses outils d'éditeur et d'annonceur, sans que l'utilisateur en ait une connaissance directe. Bien que ces données soient collectées à l'aide d'identifiants anonymes, Google a la possibilité de relier ces informations collectées aux identifiants personnels de l'utilisateur stockés sur son compte Google.

57. Il convient de souligner que la collecte passive de données d'utilisateurs de Google à partir de pages web tierces ne peut être empêchée à l'aide d'outils populaires de blocage de publicité<sup>26</sup>, car ces outils sont conçus principalement pour empêcher la présence de publicités pendant que les utilisateurs naviguent sur des pages web tierces<sup>27</sup>. La section suivante examine de plus près l'ampleur de cette collecte de données.

---

## Ce que récolte Google : revue de détail

Le temps n'est plus où il était nécessaire d'alerter sur la prédation opérée par Google et ses nombreux services sur nos données personnelles. Il est fréquent aujourd'hui d'entendre dire sur un ton fataliste : « de toute façon, ils espionnent tout »

Si beaucoup encore proclament à l'occasion « je n'ai rien à cacher » c'est moins par conviction réelle que parce que chacun en a fait l'expérience : « on ne peut rien cacher » dans le monde numérique. Depuis quelques années, les mises en garde, listes de précautions à prendre et solutions alternatives ont été largement exposées, et Framasoft parmi d'autres y a contribué.

Il manquait toutefois un travail de fond pour explorer et comprendre, une véritable étude menée suivant la démarche universitaire et qui, au-delà du jugement global approximatif, établisse les faits avec précision.

C'est à quoi s'est attelée l'équipe du professeur Douglas C. Schmidt, spécialiste depuis longtemps des systèmes logiciels, chercheur et enseignant à l'[Université Vanderbilt](#), qui livre au public une étude d'une cinquantaine de pages intitulée [Google Data Collection](#). Cette étude, qui nous semble pouvoir servir de référence, a retenu l'attention du groupe Framalang qui vous en livre ci-dessous l'*executive summary*, c'est-à-dire une sorte de résumé initial, qui en donne un aperçu programmatique.

Si vous trouvez un intérêt à cette traduction et souhaitez que Framalang vous propose la suite nous ferons de notre mieux...

**Traduction Framalang : Alain, fab, FranBAG, Goofy, jums, Khrys, Mika, Piup, serici**

## La collecte de données de Google

### Un premier aperçu

1.  
Google est la plus grosse agence de publicité numérique du monde <sup>28</sup>. Elle fournit aussi le leader des navigateurs web <sup>29</sup>, la première plateforme mobile <sup>30</sup> ainsi que le moteur de recherche le plus utilisé au monde <sup>31</sup>. La plateforme vidéo de Google, ses services de mail et de cartographie comptent 1 milliard d'utilisateurs mensuels actifs chacun <sup>32</sup>. Google utilise l'immense popularité de ses produits pour collecter des données détaillées sur le comportement des utilisateurs en ligne comme dans la vie réelle, données qu'il utilisera ensuite pour cibler ses utilisateurs avec de la publicité payante. Les revenus de Google augmentent significativement en fonction de la finesse des technologies de ciblage des données.



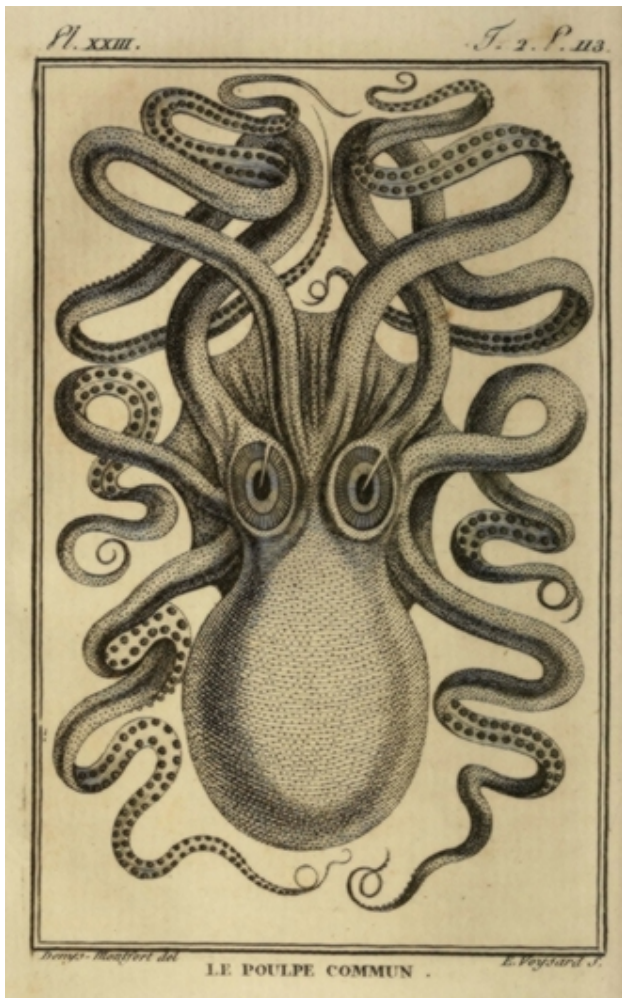
2.

Google collecte les données utilisateurs de diverses manières. Les plus évidentes sont « actives », celles dans lesquelles l'utilisateur donne directement et consciemment des informations à Google, par exemple en s'inscrivant à des applications très populaires telles que YouTube, Gmail, ou le moteur de recherche. Les voies dites « passives » utilisées par Google pour collecter des données sont plus discrètes, quand une application devient pendant son utilisation l'instrument de la collecte des données, sans que l'utilisateur en soit conscient. On trouve ces méthodes de collecte dans les plateformes (Android, Chrome), les applications (le moteur de recherche, YouTube, Maps), des outils de publication (Google Analytics, AdSense) et de publicité (AdMob, AdWords). L'étendue et l'ampleur de la collecte passive de données de Google ont été en grande partie négligées par les études antérieures sur le sujet <sup>33</sup>.

3.

Pour comprendre les données que Google collecte, cette étude s'appuie sur quatre sources clefs :

- a. Les outils Google « Mon activité » (*My Activity*) <sup>34</sup> et « Téléchargez vos données » (*Takeout*) <sup>35</sup>, qui décrivent aux utilisateurs l'information collectée lors de l'usage des outils Google.
- b. Les données interceptées lors de l'envoi aux serveurs de Google pendant l'utilisation des produits Google ou d'autres sociétés associées.
- c. Les règles concernant la vie privée (des produits Google spécifiquement ou en général).
- d. Des recherches tierces qui ont analysé les collectes de données opérées par Google.



*Histoire naturelle, générale et particulière, des mollusques, animaux sans vertèbres et à sang blanc. T.2. Paris, L'Imprimerie de F. Dufart, An X-XIII [1802-1805]. [biodiversitylibrary.org/page/35755415](http://biodiversitylibrary.org/page/35755415)*

4.

Au travers de la combinaison des sources ci-dessus, cette étude montre une vue globale et exhaustive de l'approche de Google concernant la collecte des données et aborde en profondeur certains types d'informations collectées auprès des utilisateurs et utilisatrices.

Cette étude met en avant les éléments clés suivants :

a. Dans une journée d'utilisation typique, Google en apprend

énormément sur les intérêts personnels de ses utilisateurs. Dans ce scénario d'une journée « classique », où un utilisateur réel avec un compte Google et un téléphone Android (avec une nouvelle carte SIM) suit sa routine quotidienne, Google collecte des données tout au long des différentes activités, comme la localisation, les trajets empruntés, les articles achetés et la musique écoutée. De manière assez surprenante, Google collecte ou infère plus de deux tiers des informations via des techniques passives. Au bout du compte, Google a identifié les intérêts des utilisateurs avec une précision remarquable.

b. Android joue un rôle majeur dans la collecte des données pour Google, avec plus de 2 milliards d'utilisateurs actifs mensuels dans le monde <sup>36</sup>. Alors que le système d'exploitation Android est utilisé par des fabricants d'équipement d'origine (FEO) partout dans le monde, il est étroitement connecté à l'écosystème Google via le service Google Play. Android aide Google à récolter des informations personnelles sur les utilisateurs (nom, numéro de téléphone, date de naissance, code postal et dans beaucoup de cas le numéro de carte bancaire), les activités réalisées sur le téléphone (applications utilisées, sites web consultés) et les coordonnées de géolocalisation. En coulisses, Android envoie fréquemment la localisation de l'utilisateur ainsi que des informations sur l'appareil lui-même, comme sur l'utilisation des applications, les rapports de bugs, la configuration de l'appareil, les sauvegardes et différents identifiants relatifs à l'appareil.

c. Le navigateur Chrome aide Google à collecter des données utilisateurs depuis à la fois le téléphone et l'ordinateur de bureau, grâce à quelque 2 milliards d'installations dans le monde <sup>37</sup>. Le navigateur Chrome collecte des informations personnelles (comme lorsqu'un utilisateur remplit un formulaire en ligne) et les envoie à Google via le processus de synchronisation. Il liste aussi les pages visitées et

envoie les données de géolocalisation à Google.

d. Android comme Chrome envoient des données à Google même en l'absence de toute interaction de l'utilisateur. Nos expériences montrent qu'un téléphone Android dormant et stationnaire (avec Chrome actif en arrière-plan) a communiqué des informations de localisation à Google 340 fois pendant une période de 24 heures, soit en moyenne 14 communications de données par heure. En fait, les informations de localisation représentent 35 % de l'échantillon complet de données envoyés à Google. À l'opposé, une expérience similaire a montré que sur un appareil iOS d'Apple avec Safari (où ni Android ni Chrome n'étaient utilisés), Google ne pouvait pas collecter de données notables (localisation ou autres) en absence d'interaction de l'utilisateur avec l'appareil.

e. Une fois qu'un utilisateur ou une utilisatrice commence à interagir avec un téléphone Android (par exemple, se déplace, visite des pages web, utilise des applications), les communications passives vers les domaines de serveurs Google augmentent considérablement, même dans les cas où l'on n'a pas utilisé d'applications Google majeures (c.-à-d. ni recherche Google, ni YouTube, pas de Gmail ni Google Maps). Cette augmentation s'explique en grande partie par l'activité sur les données de l'éditeur et de l'annonceur de Google (Google Analytics, DoubleClick, AdWords) <sup>38</sup>. Ces données représentaient 46 % de l'ensemble des requêtes aux serveurs Google depuis le téléphone Android. Google a collecté la localisation à un taux 1,4 fois supérieur par rapport à l'expérience du téléphone fixe sans interaction avec l'utilisateur. En termes d'amplitude, les serveurs de Google ont communiqué 11,6 Mo de données par jour (ou 0,35 Go / mois) avec l'appareil Android. Cette expérience suggère que même si un utilisateur n'interagit avec aucune application phare de Google, Google est toujours en mesure de recueillir beaucoup d'informations par l'entremise de ses produits d'annonce et d'éditeur.

f. Si un utilisateur d'appareil sous iOS décide de renoncer à l'usage de tout produit Google (c'est-à-dire sans Android, ni Chrome, ni applications Google) et visite exclusivement des pages web non-Google, le nombre de fois où les données sont communiquées aux serveurs de Google demeure encore étonnamment élevé. Cette communication est menée exclusivement par des services de l'annonceur/éditeur. Le nombre d'appels de ces services Google à partir d'un appareil iOS est similaire à ceux passés par un appareil Android. Dans notre expérience, la quantité totale de données communiquées aux serveurs Google à partir d'un appareil iOS est environ la moitié de ce qui est envoyé à partir d'un appareil Android.

g. Les identificateurs publicitaires (qui sont censés être « anonymisés » et collectent des données sur l'activité des applications et les visites des pages web tierces) peuvent être associés à l'identité d'un utilisateur ou utilisatrice de Google. Cela se produit par le transfert des informations d'identification depuis l'appareil Android vers les serveurs de Google. De même, le cookie ID DoubleClick (qui piste les activités des utilisateurs et utilisatrices sur les pages web d'un tiers) constitue un autre identificateur censé être anonymisé que Google peut associer à celui d'un compte personnel Google, si l'utilisateur accède à une application Google avec le navigateur déjà utilisé pour aller sur la page web externe. En définitive, nos conclusions sont que Google a la possibilité de connecter les données anonymes collectées par des moyens passifs avec les données personnelles de l'utilisateur.