



Ce que collecte Google

Août 2018



Cette étude a été menée par le professeur Douglas C. Schmidt, enseignant en informatique à l'université Vanderbilt, et par son équipe.

DCN remercie le professeur Schmidt d'en autoriser la diffusion. Nous l'offrons donc au public avec son autorisation.

La version originale en anglais est disponible à l'adresse :

<https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf>

La traduction française, publiée initialement sur le Framablog, est due à l'équipe Framalang (crédits en annexe).

Table des matières

I. Un premier aperçu.....	4
II. Une journée dans la vie d'une utilisatrice de Google.....	8
III. La collecte des données par les plateformes Android et Chrome.....	12
A. Collecte d'informations personnelles et de données d'activité.....	12
B. Collecte des données de localisation de l'utilisateur.....	13
C. Une évaluation de la collecte passive de données par Google via Android et Chrome.....	16
IV. Collecte de données par les outils des annonceurs et des diffuseurs.....	19
A. Google Analytics et DoubleClick.....	21
B. AdSense, AdWords et AdMob.....	22
C. Association de données recueillies passivement et d'informations à caractère personnel.....	23
V. Quantité de données collectées lors d'une utilisation minimale des produits Google.....	28
VI. Données collectées par les applications clés de Google destinées aux particuliers.....	31
A. Recherche.....	31
B. YouTube.....	32
C. Maps.....	34
D. Gmail.....	35
VII. Des produits avec un haut potentiel futur d'agrégation de données.....	37
A. Pages optimisées pour les mobiles (AMP).....	37
B. Google Assistant.....	39
C. Photos.....	40
D. Chromebook.....	41
E. Google Pay.....	42
F. Données d'utilisateurs collectées auprès de fournisseurs de données tiers.....	42
VIII. CONCLUSION.....	43
IX. ANNEXE.....	44
A. Caractérisation des collections de données actives vs passives à partir d'une journée-type d'un utilisateur.....	44
B. Liste des produits Google.....	44
C. Collecte de données via d'autres produits majeurs de Google.....	45
D. Méthode du suivi de trafic selon la localisation.....	52
E. Séquence d'authentification de l'inscription à Google.....	56
F. Profil d'utilisation pour les expériences de collecte de données du mobile.....	56
G. Articles publiés sur les pratiques de collecte de données mises en œuvre par Google.....	57
H. Clarifications.....	58
I. À propos de l'auteur.....	58
J. Crédits de traduction.....	59

I. Un premier aperçu

1. Google est la plus grosse agence de publicité numérique du monde¹. Elle fournit aussi le leader des navigateurs web², la première plateforme mobile³ ainsi que le moteur de recherche le plus utilisé au monde⁴. La plateforme vidéo de Google, ses services de courriel et de cartographie comptent 1 milliard d'utilisateurs mensuels actifs chacun⁵. 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. Une liste des études antérieures ou nouvelles sur la collecte de données par Google figure dans l'appendice du présent document.

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

- a) Les outils Google « Mon activité » ("[My Activity](#)") et « Téléchargez vos données » ("[Takeout](#)"), 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.

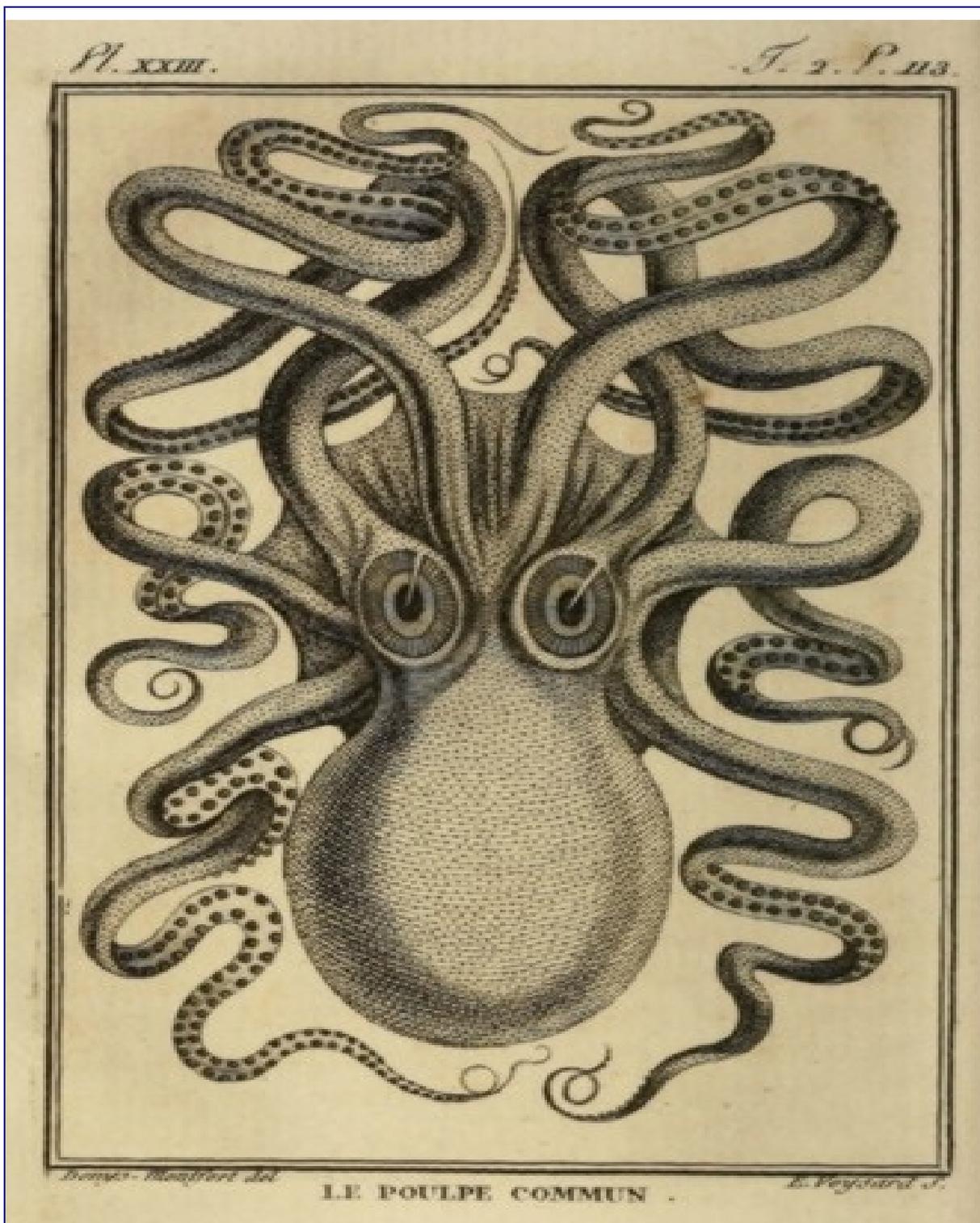
1 "[Google and Facebook tighten grip on US digital ad market](#)," eMarketer, 21 septembre 2017

2 "[Market share or leading internet browsers in the United States and worldwide as of February 2018](#)," Statista, février 2018

3 "[Global OS market share in sales to end users from 1st quarter 2009 to 2nd quarter 2017](#)," Statista, août 2017

4 "[Worldwide desktop market share of leading search engines from January 2010 to October 2017](#)," Statista, février 2018

5 [Google 10 K filings with the SEC](#), 2017



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].

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 monde⁶. 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 monde⁷. 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) Google a récemment renommé sa régie publicitaire AdWords en

6 Dave Burke, "[Android : celebrating a big milestone together with you](#)," Google, 17 mai 2017

7 Frederic Lardinois, "[Google says there are now 2 billion active Chrome installs](#)," TechCrunch, 10 novembre 2016

“Google Ads” et DoubleClick devient “Google Ad Manager”. 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.

II. Une journée dans la vie d'une utilisatrice de Google



Photo susanjanegolding (CC BY 2.0)

5. Afin d'illustrer la multitude des interactions entre Google et un individu, ainsi que l'étendue des informations collectées lors de ces interactions, nous avons réalisé une expérience dans laquelle un chercheur utilise un périphérique Android⁸ pendant les activités d'une journée.

Afin d'éviter que des informations d'un utilisateur précédent ne soient associées au téléphone mobile, celui-ci a été réinitialisé aux valeurs d'usine⁹ et configuré comme un téléphone neuf¹⁰. Un nouveau compte Google a été créé (nom d'utilisatrice : « Jane »), afin que Google n'ait pas de connaissances antérieures sur cette utilisatrice et qu'il n'ait pas associé de centres d'intérêts publicitaires à son compte. Le chercheur a passé une journée normale en utilisant son téléphone avec son nouveau compte Google.

6. Les données collectées par Google ont été relevées par deux outils fournis par Google : « Mon activité »¹¹ et « Télécharger vos données »¹². L'outil « Mon activité » montre les données collectées

8 Téléphone LG X Power avec la version 6.0 d'Android

9 La réinitialisation aux valeurs d'usine efface toutes les informations du stockage interne du téléphone. Cela concerne les données de connexion aux services Google et des autres comptes, les informations système, les données applicatives et les configurations, toutes les applications téléchargées, les licences DRM, la musique, les images, les documents, les sauvegardes et toutes les données d'utilisation.

10 Les chercheurs ont utilisé un téléphone LG X Power remis à sa configuration d'usine et équipé d'une carte SIM neuve pour garantir qu'aucune donnée n'était stockée sur le téléphone et que le numéro de téléphone ne pourrait pas être rattaché à un usage précédent.

11 "My Activity," Google, disponible à l'adresse <https://myactivity.google.com/myactivity>

12 "Download your data," Google, disponible à l'adresse <https://takeout.google.com/settings/takeout?pli=1>

par Google grâce à toute activité liée aux recherches, lors de l'utilisation des applications Google (i.e Youtube, Google Maps, Google assistant), par les visites sur des pages web tierces (lorsqu'on est connecté à Chrome), et les clics sur les publicités. L'outil « Télécharger vos données » fournit des informations structurées concernant l'historique de toutes les données collectées par les applications Google (en fait elles contiennent tous les anciens courriels sur Gmail, toutes les recherches, l'ensemble des localisations et les vidéos YouTube consultées). Nous avons synthétisé les données collectées et nous les avons utilisées pour représenter les informations sur les événements clés dans l'infographie ci-dessous : « Un jour dans la vie » de l'utilisateur "Jane".

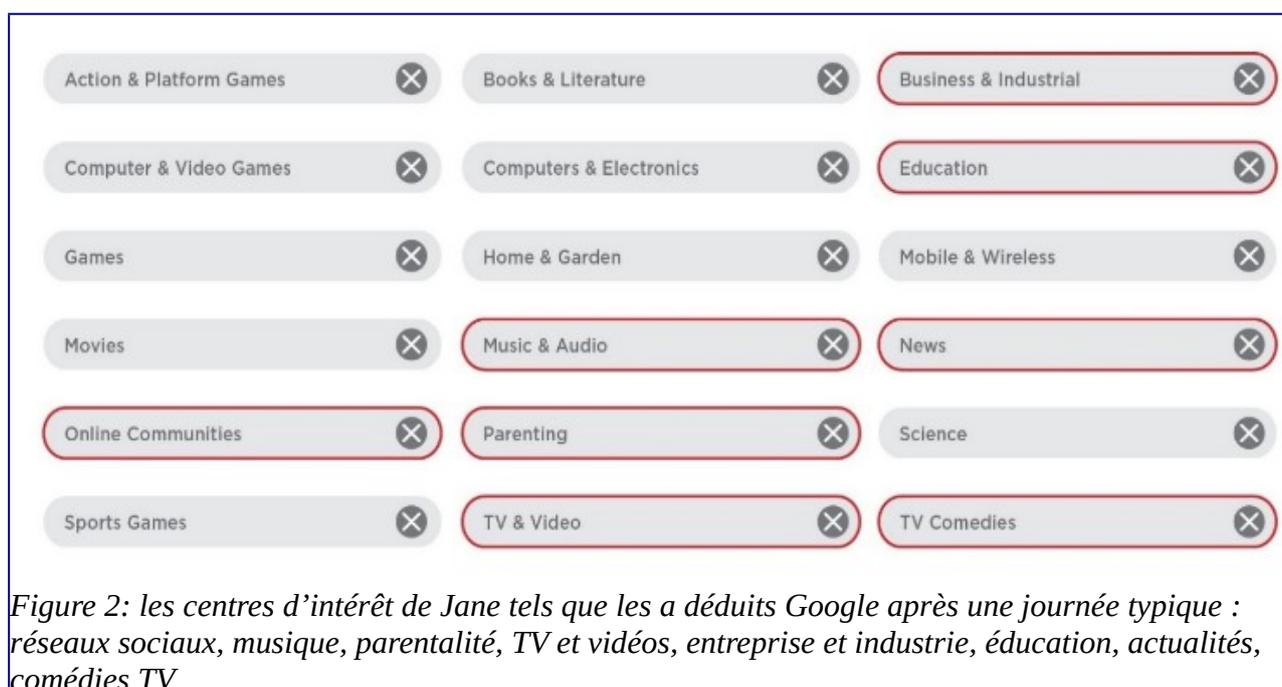


7. Pour l'activité détaillée dans la figure 1, mais également dans le reste du document, les données collectées sont classées en deux grands groupes : actives et passives. Les données actives sont des données échangées directement entre l'utilisateur et un produit Google, là où les données passives sont définies comme une information transmise en arrière-plan sans notification évidente pour l'utilisateur. Par exemple, une collecte active de données est déclenchée lorsque Jane saisit un mot-

clé dans l’outil de recherche et que cette requête est enregistrée par Google. Un exemple de collecte de données passives est l’envoi de la localisation de Jane à Google suite à l’enregistrement d’une requête.

8. L’analyse des points de contact clés durant une journée normale dans la vie de Jane suggère que le nombre de données passives transmises est deux fois plus grand que le nombre de données actives (une décomposition détaillée des caractéristiques des données actives et passives est fournie dans le tableau qui figure en appendice.

9. Google analyse les données collectées pour déterminer les centres d’intérêt des utilisateurs et utilisatrices, qu’il utilise ensuite pour les cibler avec des publicités adaptées. Par exemple, Google fournit une liste d’intérêts qu’il a déduits de l’activité d’un utilisateur, que l’on peut consulter sur la page « Les sujets qui vous intéressent » de la page de « Personnalisation des annonces » de Google¹³. La figure 2 ci-dessous montre la liste que Google a associée avec le compte de Jane après une journée d’activité. Au total, Google a attribué 18 centres d’intérêts à Jane, dont 8 (entourés par une bordure rouge) qui correspondent précisément aux utilisations et activités de Jane.¹⁴



10. Bien que les outils « Mon activité » et « Télécharger mes données » soient utiles pour estimer la quantité de données actives collectées lors des interactions d’un utilisateur avec les produits Google, ils ne dessinent pas une image complète de l’ampleur et de l’échelle de la collecte de données de Google. Comprendre cela requiert un passage en revue détaillé des clauses d’utilisation des produits en ce qui concerne la vie privée mais également l’analyse du trafic de données envoyé aux serveurs de Google pendant une session d’utilisation par un utilisateur de ces services. Les résultats de cette analyse sont présentés plus loin dans ce rapport.

13 “Ads personalization,” Google, consulté le 15 août 2018, disponible à l’adresse <https://adssettings.google.com/authenticated>

14 La raison de la présence d’autres centres d’intérêt sans aucune connexion avec les activités de Jane durant la journée n’est pas claire. Google utilise sans doute des résultats d’analyses concernant des historiques d’utilisateurs ayant des intérêts similaires.

III. La collecte des données par les plateformes Android et Chrome

11. Android et Chrome sont les plateformes clés de Google qui facilitent la collecte massive de données des utilisateurs en raison de leur grande portée et fréquence d'utilisation. En janvier 2018, Android détenait 53 % du marché américain des systèmes d'exploitation mobiles (iOS d'Apple en détenait 45 %) ¹⁵ et, en mai 2017, il y avait plus de 2 milliards d'appareils Android actifs par mois dans le monde. ¹⁶

12. Le navigateur Chrome de Google représentait plus de 60 % de l'utilisation mondiale de navigateurs Internet avec plus d'un milliard d'utilisateurs actifs par mois, comme l'indiquait le rapport Q4 10 K de 2017 ¹⁷. Les deux plateformes facilitent l'usage de contenus de Google et de tiers (par ex. applications et sites tiers) et fournissent donc à Google un accès à un large éventail d'informations personnelles, d'activité web, et de localisation.

A. Collecte d'informations personnelles et de données d'activité

13. Pour télécharger et utiliser des applications depuis le Google Play Store sur un appareil Android, un utilisateur doit posséder (ou créer) un compte Google, qui devient une passerelle clé par laquelle Google collecte ses informations personnelles, ce qui comporte son nom d'utilisateur, son adresse de messagerie et son numéro de téléphone. Si un utilisateur s'inscrit à des services comme Google Pay ¹⁸, Android collecte également les données de la carte bancaire, le code postal et la date de naissance de l'utilisateur. Toutes ces données font alors partie des informations personnelles de l'utilisateur associées à son compte Google.

14. Alors que Chrome n'oblige pas le partage d'informations personnelles supplémentaires recueillies auprès des utilisateurs, il a la possibilité de récupérer de telles informations. Par exemple, Chrome collecte toute une série d'informations personnelles avec la fonctionnalité de remplissage automatique des formulaires, qui incluent typiquement le nom d'utilisateur, l'adresse, le numéro de téléphone, l'identifiant de connexion et les mots de passe ¹⁹. Chrome stocke les informations saisies dans les formulaires sur le disque dur de l'utilisateur. Cependant, si l'utilisateur se connecte à Chrome avec un compte Google et active la fonctionnalité de synchronisation, ces informations sont envoyées et stockées sur les serveurs de Google. Chrome pourrait également apprendre la ou les langues que parle la personne avec sa fonctionnalité de traduction, activée par défaut ²⁰.

15 "Subscriber share held by smartphone operating systems in the United States from 2012 to 2018" (« Part des abonnés détenue par les systèmes d'exploitation de smartphones aux États-Unis de 2012 à 2018 »), Statista, mai 2018, disponible à l'adresse <https://www.statista.com/statistics/266572/market-share-held-by-smartphone-platforms-in-the-united-states/>

16 Dave Burke, "Android : celebrating a big milestone together with you" (« Android : célébrer un grand tournant avec vous »), Google, 17 mai 2017, disponible à l'adresse <https://www.blog.google/products/android/2bn-milestone/>

17 Formulaire 10 K de Google pour la SEC ("Securities and Exchange Commission")

18 "Google Chrome privacy whitepaper" (« Livre blanc de Google Chrome sur la confidentialité »), Google, 6 mars 2018, disponible à l'adresse <https://www.google.com/chrome/privacy/whitepaper.html#payments>

19 "Google Chrome privacy whitepaper" (« Livre blanc de Google Chrome sur la confidentialité »), Google, 6 mars 2018, disponible à l'adresse <https://www.google.com/chrome/privacy/whitepaper.html#autofill>

20 "Google Chrome privacy whitepaper" (« Livre blanc de Google Chrome sur la confidentialité »), Google, 6 mars 2018, disponible à l'adresse <https://www.google.com/chrome/privacy/whitepaper.html#translate>

15. En plus des données personnelles, Chrome et Android envoient tous deux à Google des informations concernant les activités de navigation et l'emploi d'applications mobiles, respectivement. Chaque visite de page internet est automatiquement traquée et collectée par Google si l'utilisateur a un compte Chrome. Chrome collecte également son historique de navigation, ses mots de passe, les permissions particulières selon les sites web, les cookies, l'historique de téléchargement et les données relatives aux extensions²¹.

16. Android envoie des mises à jour régulières aux serveurs de Google, ce qui comprend le type d'appareil, le nom de l'opérateur, les rapports de bug et des informations sur les applications installées²². Il avertit également Google chaque fois qu'une application est ouverte sur le téléphone (ex. Google sait quand un utilisateur d'Android ouvre son application Uber).

B. Collecte des données de localisation de l'utilisateur

17. Android et Chrome collectent méticuleusement la localisation et les mouvements de l'utilisateur en utilisant une variété de sources, représentées sur la figure 3. Par exemple, un accès à la « localisation approximative » peut être réalisé en utilisant les coordonnées GPS sur un téléphone Android ou avec l'adresse IP sur un ordinateur. La précision de la localisation peut être améliorée (« localisation précise ») avec l'usage des identifiants des antennes cellulaires environnantes ou en scannant les BSSID (*Basic Service Set Identifiers*), identifiants assignés de manière unique aux puces radio des points d'accès Wi-Fi présents aux alentours²³. Les téléphones Android peuvent aussi utiliser les informations des balises Bluetooth enregistrées dans l'API Proximity Beacon de Google²⁴. Ces balises non seulement fournissent les coordonnées de géolocalisation de l'utilisateur, mais pourraient aussi indiquer à quel étage exact il se trouve dans un immeuble²⁵.

21 "Google Chrome Privacy Notice", (« Déclaration de confidentialité de Google Chrome »), Google, 6 mars 2018, disponible à l'adresse <https://www.google.com/intl/en/chrome/browser/privacy>

22 <https://policies.google.com/privacy?hl=en&gl=us#infocollect>

23 Pour mieux comprendre comment les données de localisation sont transmises aux serveurs Google, nos chercheurs ont analysé le trafic de données d'un téléphone mobile d'un utilisateur en mouvement, en appliquant la méthode décrite à l'annexe VIII.C.

24 "Google beacon platform, proximity beacon API", Google, visité pour la dernière fois le 10 août 2018, disponible à l'adresse <https://developers.google.com/beacons/proximity/guides>

25 "Google beacon platform, proximity beacon API", Google, visité pour la dernière fois le 10 août 2018, disponible à l'adresse <https://developers.google.com/beacons/proximity/guides>

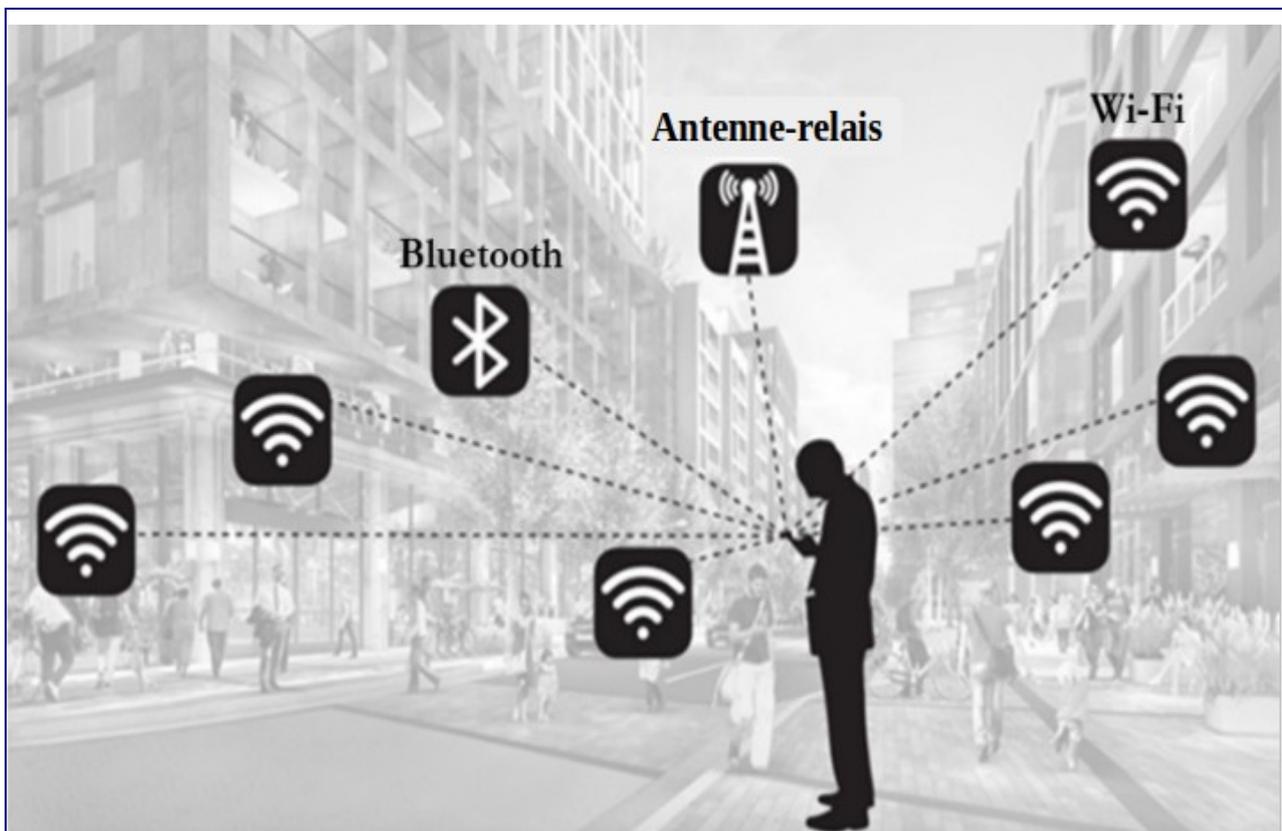


Figure 3: *Android et Chrome utilisent diverses manières de localiser l'utilisateur d'un téléphone.*

18. Il est difficile pour un utilisateur de téléphone Android de refuser le traçage de sa localisation. Par exemple, sur un appareil Android, même si un utilisateur désactive le Wi-Fi, la localisation est toujours suivie par son signal Wi-Fi. Pour éviter un tel traçage, le scan Wi-Fi doit être explicitement désactivé par une autre action de l'utilisateur, comme montré sur la figure 4.

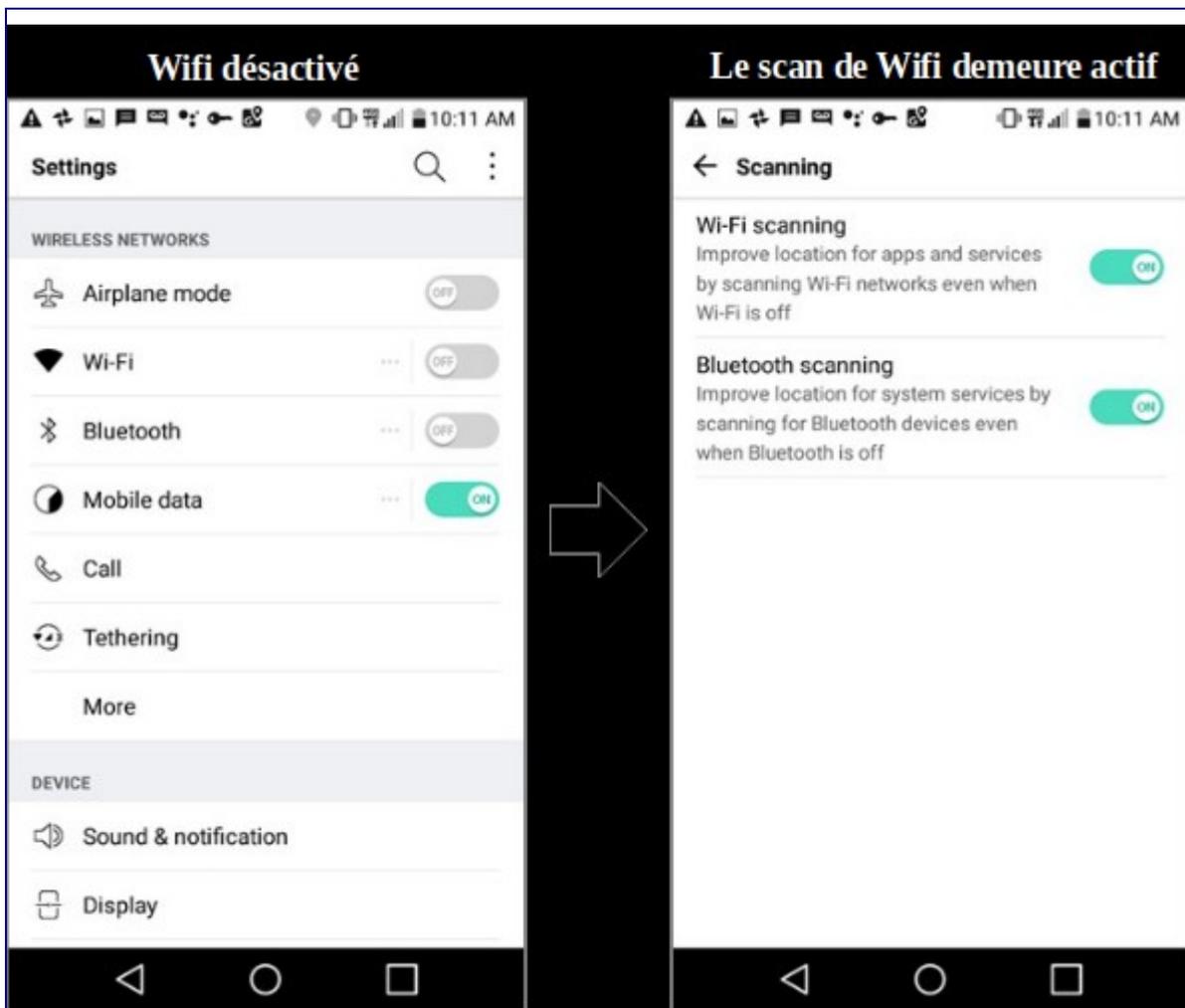


Figure 4: Android collecte des données même si le Wi-Fi est éteint par l'utilisateur

19. L'omniprésence de points d'accès Wi-Fi a rendu le traçage de localisation assez fréquent. Par exemple, durant une courte promenade de 15 minutes autour d'une résidence, un appareil Android a envoyé neuf requêtes de localisation à Google. Les requêtes contenaient au total environ 100 BSSID de points d'accès Wi-Fi publics et privés.

20. Google peut vérifier avec un haut degré de confiance si un utilisateur est immobile, s'il marche, court, fait du vélo, ou voyage en train ou en car. Il y parvient grâce au traçage à intervalles de temps réguliers de la localisation d'un utilisateur Android, combiné aux données des capteurs embarqués (comme l'accéléromètre) sur les téléphones mobiles. La figure 5 montre un exemple de telles données communiquées aux serveurs de Google pendant que l'utilisateur marchait.

```
"activityReadings": [
  {
    "activities": [
      {
        "confidence": 99,
        "type": "onFoot"
      },
      {
        "confidence": 99,
        "type": "walking"
      },
      {
        "confidence": 1,
        "type": "unknown"
      }
    ],
    "timestampMs": 1527095517507
  },
]
```

Figure 5: capture d'écran d'un envoi de localisation d'utilisateur à Google.

C. Une évaluation de la collecte passive de données par Google via Android et Chrome

21. Les données actives que les plateformes Android ou Chrome collectent et envoient à Google à la suite des activités des utilisateurs sur ces plateformes peuvent être évaluées à l'aide des outils *MyActivity* et *Takeout*. Les données passives recueillies par ces plateformes, qui vont au-delà des données de localisation et qui restent relativement méconnues des utilisateurs, présentent cependant un intérêt potentiellement plus grand. Afin d'évaluer plus en détail le type et la fréquence de cette collecte, une expérience a été menée pour surveiller les données relatives au trafic envoyées à Google par les téléphones mobiles (Android et iPhone) en utilisant la méthode décrite dans la section IX.D de l'annexe. À titre de comparaison, cette expérience comprenait également l'analyse des données envoyées à Apple via un appareil iPhone.

22. Pour des raisons de simplicité, les téléphones sont restés stationnaires, sans aucune interaction avec l'utilisateur. Sur le téléphone Android, une seule session de navigateur Chrome restait active en arrière-plan, tandis que sur l'iPhone, le navigateur Safari était utilisé. Cette configuration a permis une analyse systématique de la collecte de fond que Google effectue uniquement via Android et Chrome, ainsi que de la collecte qui se produit en l'absence de ceux-ci (c'est-à-dire à partir d'un appareil iPhone), sans aucune demande de collecte supplémentaire générée par d'autres produits et applications (par exemple YouTube, Gmail ou utilisation d'applications).

23. La figure 6 présente un résumé des résultats obtenus dans le cadre de cette expérience. L'axe des abscisses indique le nombre de fois où les téléphones ont communiqué avec les serveurs Google (ou Apple), tandis que l'axe des ordonnées indique le type de téléphone (Android ou iPhone) et le type de domaine de serveur (Google ou Apple) avec lequel les paquets de données ont été échangés par les téléphones. La légende en couleur décrit la catégorisation générale du type de demandes de données identifiées par l'adresse de domaine du serveur. Une liste complète des adresses de domaine appartenant à chaque catégorie figure dans le tableau 5 de la section IX.D de l'annexe.

24. Au cours d'une période de 24 heures, l'appareil Android a communiqué environ 900 échantillons de données à une série de terminaux de serveur Google. Parmi ceux-ci, environ 35 % (soit environ 14 par heure) étaient liés à la localisation. Les domaines publicitaires de Google n'ont reçu que 3 % du trafic, ce qui est principalement dû au fait que le navigateur mobile n'a pas été utilisé activement pendant la période de collecte. Le reste (62 %) des communications avec les domaines de serveurs Google se répartissaient grosso modo entre les demandes adressées au magasin d'applications Google Play, les téléchargements par Android de données relatives aux périphériques (tels que les rapports de crash et les autorisations de périphériques), et d'autres données, principalement de la catégorie des appels et actualisations de fond des services Google.

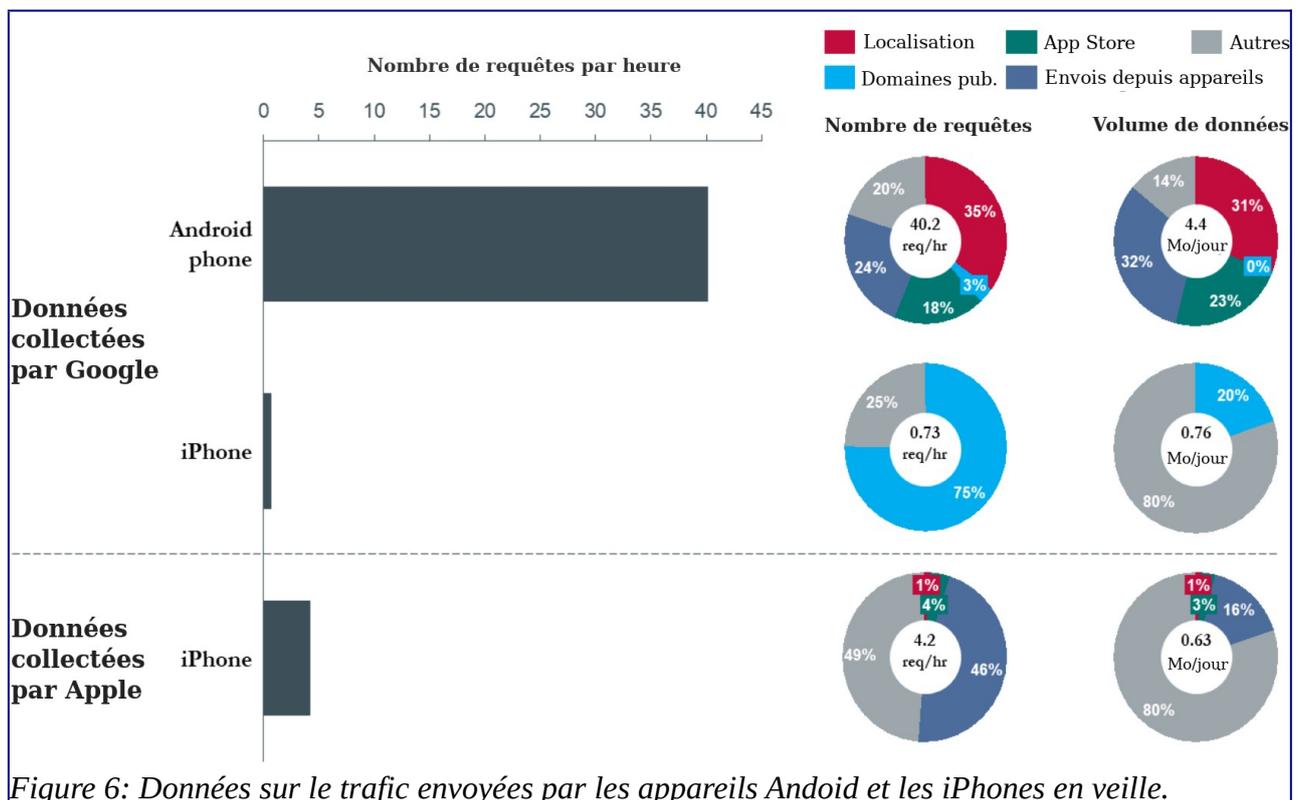


Figure 6: Données sur le trafic envoyées par les appareils Android et les iPhones en veille.

25. La figure 6 montre que l'appareil iPhone communiquait avec les domaines Google à une fréquence inférieure de plus d'un ordre de grandeur (50 fois) à celle de l'appareil Android, et que Google n'a recueilli aucune donnée de localisation utilisateur pendant la période d'expérience de 24 heures via iPhone. Ce résultat souligne le fait que les plateformes Android et Chrome jouent un rôle important dans la collecte de données de Google.

26. De plus, les communications de l'appareil iPhone avec les serveurs d'Apple étaient 10 fois moins fréquentes que les communications de l'appareil Android avec Google. Les données de localisation

ne représentaient qu'une très faible fraction (1 %) des données nettes envoyées aux serveurs Apple à partir de l'iPhone, Apple recevant en moyenne une fois par jour des communications liées à la localisation.

27. En termes d'amplitude, les téléphones Android communiquaient 4,4 Mo de données par jour (130 Mo par mois) avec les serveurs Google, soit 6 fois plus que ce que les serveurs Google communiquaient à travers l'appareil iPhone.

28. Pour rappel, cette expérience a été réalisée à l'aide d'un téléphone stationnaire, sans interaction avec l'utilisateur. Lorsqu'un utilisateur commence à bouger et à interagir avec son téléphone, la fréquence des communications avec les serveurs de Google augmente considérablement. La section V du présent rapport résume les résultats d'une telle expérience.

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 aux 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 AdMob²⁶, plus d'un million d'annonceurs utilisent AdWords²⁷, plus de 15 millions de sites internet utilisent AdSense²⁸ et plus de 30 millions de sites utilisent Google Analytics²⁹.

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 produits³⁰. 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.

26 "AdMob by Google", Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://www.google.com/admob/>

27 "Hear from our happy customers" Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://adwords.google.com/home/resources/success-stories/>

28 "Websites using Google AdSense", BuiltWith, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://trends.builtwith.com/websitelist/Google-AdSense>

29 "Google Analytics usage statistics", BuiltWith, Avril 2018, disponible à l'adresse <https://trends.builtwith.com/analytics/Google-Analytics>

30 Garrett Sloane, "Google to retire Doubleclick and AdWords names in a rebrand of its ad business", Ad Age, disponible à l'adresse <http://adage.com/article/digital/google-waves-goodbye-doubleclick-ad-business-evolves/314046/>

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 de manière imperceptible 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.

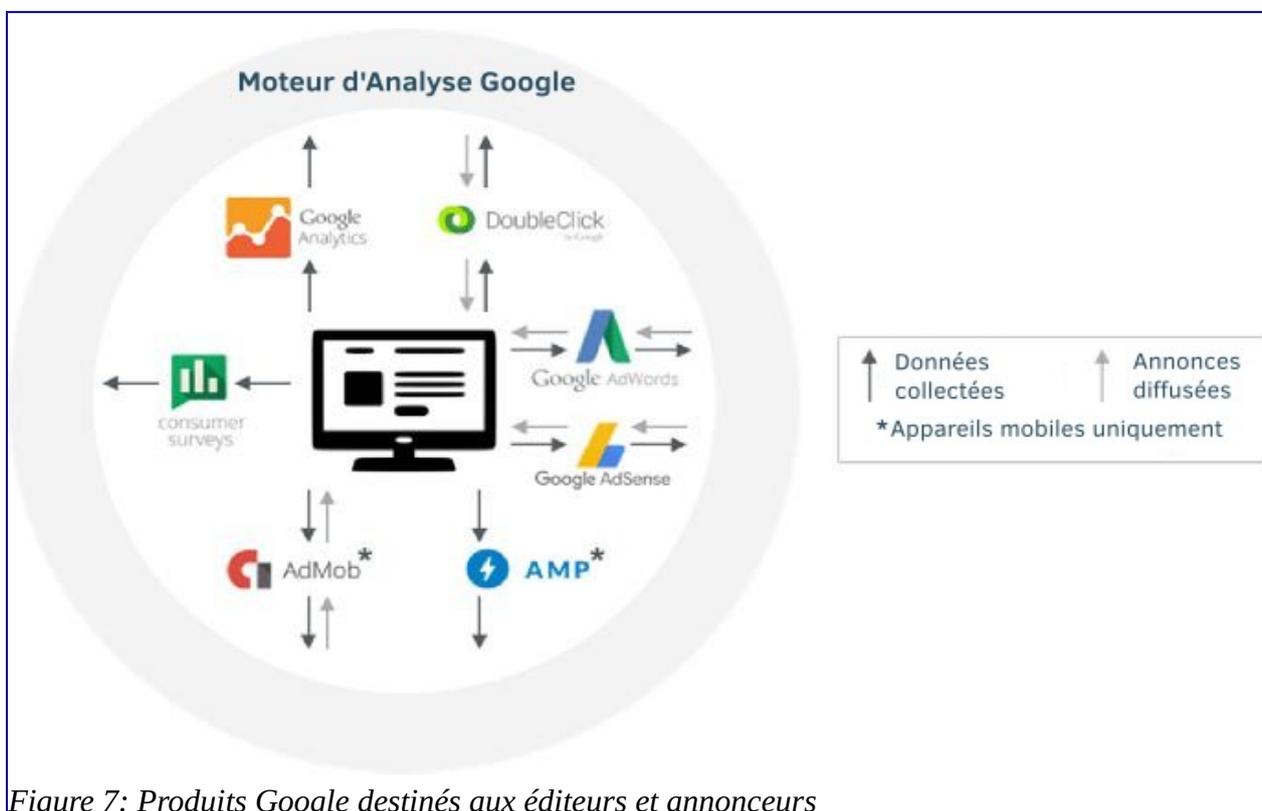


Figure 7: Produits Google destinés aux éditeurs et annonceurs

31

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)³². 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.

31 "Our products", Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://www.google.com/about/products/>

32 "Best practices for unique identifiers", Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://developer.android.com/training/articles/user-data-ids>

Tableau 1 : Identificateurs transmis à Google

Identificateur	Type	Description
GAID/IDFA	Semi-permanent	Chaîne 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	Chaîne 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 visités³³. Les cookies DoubleClick sont associés à plus de 1,6 million de sites Web³⁴.

35. GA utilise de petits segments de code de traçage (appelés « balises de page ») intégrés dans le code HTML d'un site Web³⁵. 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.

33 "Google Analytics usage statistics", BuiltWith, April 2018, disponible à l'adresse <https://trends.builtwith.com/analytics/Google-Analytics>

34 "DoubleClick market share", Datanyze, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://www.datanyze.com/market-share/ad-exchanges/doubleclick-market-share>

35 GA ou d'autres balises peuvent aussi être implémentées avec l'outil Google Tag Manage (GTM) sans modifier la fonctionnalité de balisage de page

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)³⁶ 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 navigateur³⁷

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 tiers³⁸ 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 unique³⁹ 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 sponsorisées⁴⁰ 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 internautes⁴¹ 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'annonce⁴². 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 achat⁴³.

36 "Cookies and user identification", Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://developers.google.com/analytics/devguides/collection/analyticsjs/cookies-user-id>

37 "Cookies and user identification", Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://developers.google.com/analytics/devguides/collection/analyticsjs/cookies-user-id>

38 "DoubleClick search help", Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://support.google.com/ds/answer/7298761?hl=en>

39 "DoubleClick search help", Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse https://support.google.com/ds/answer/2903014?hl=en&ref_topic=6054260

40 "Websites using Google AdSense", BuiltWith, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://trends.builtwith.com/websitelist/Google-AdSense>

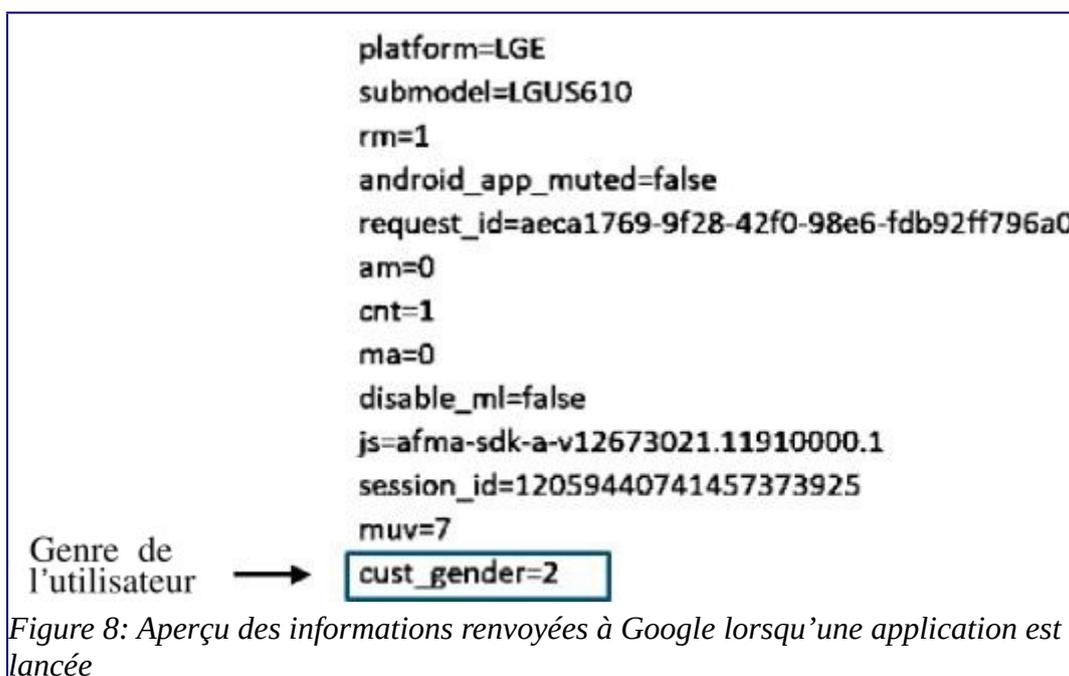
41 "Google Ads Help", Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://support.google.com/google-ads/answer/2404191?hl=en>

42 AdSense help, privacy and security", Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://support.google.com/adsense/answer/9897?hl=en>

43 "Evaluating ad performance on the Search Network", Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://support.google.com/adwords/answer/2404037?hl=en> ; "About conversion tracking", Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://support.google.com/adwords/answer/1722022?hl=en>

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 »⁴⁴ 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 (Illustration 8), et comme on peut trouver confirmation sur les propres pages web des développeurs de Google⁴⁵, 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.



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

⁴⁴ Il en va de même pour les ordinateurs de bureau, avec lesquels les cookies ne sont pas partagés d'un navigateur à l'autre.

⁴⁵ Google APIs for Android, Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://developers.google.com/android/reference/com/google/android/gms/ads/doubleclick/PublisherAdRequest.Builder>

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.

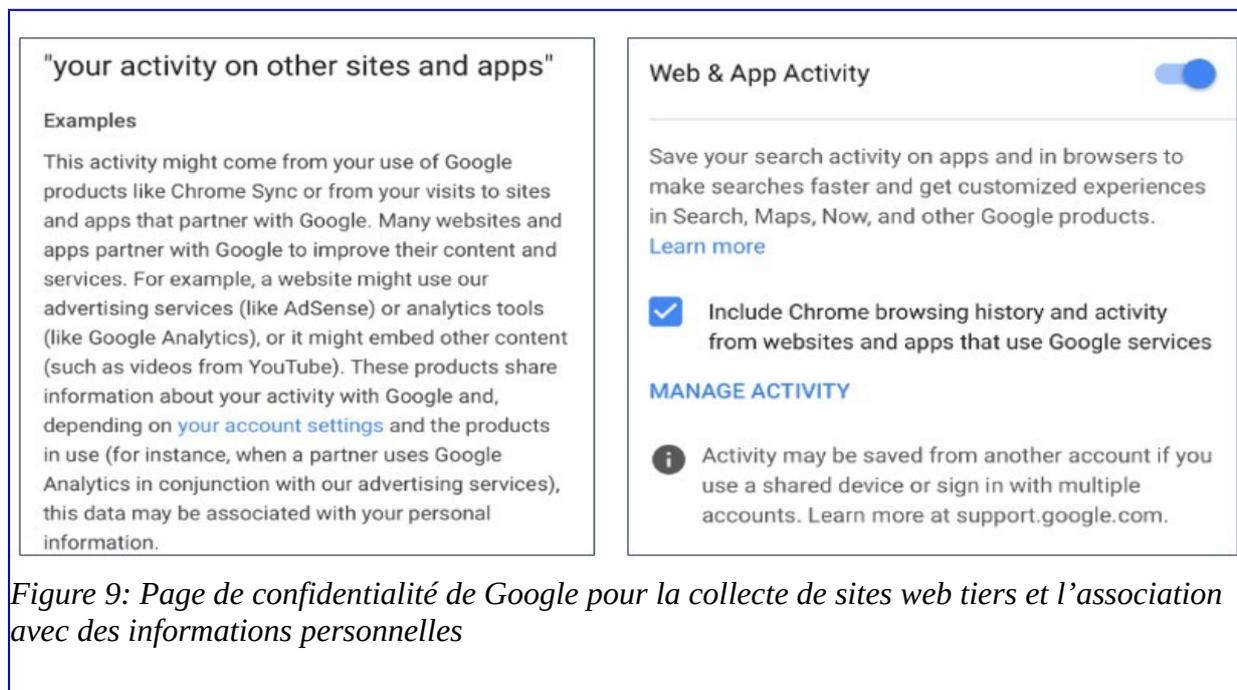


Figure 9: Page de confidentialité de Google pour la collecte de sites web tiers et l'association avec des informations personnelles

46 et 47

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.

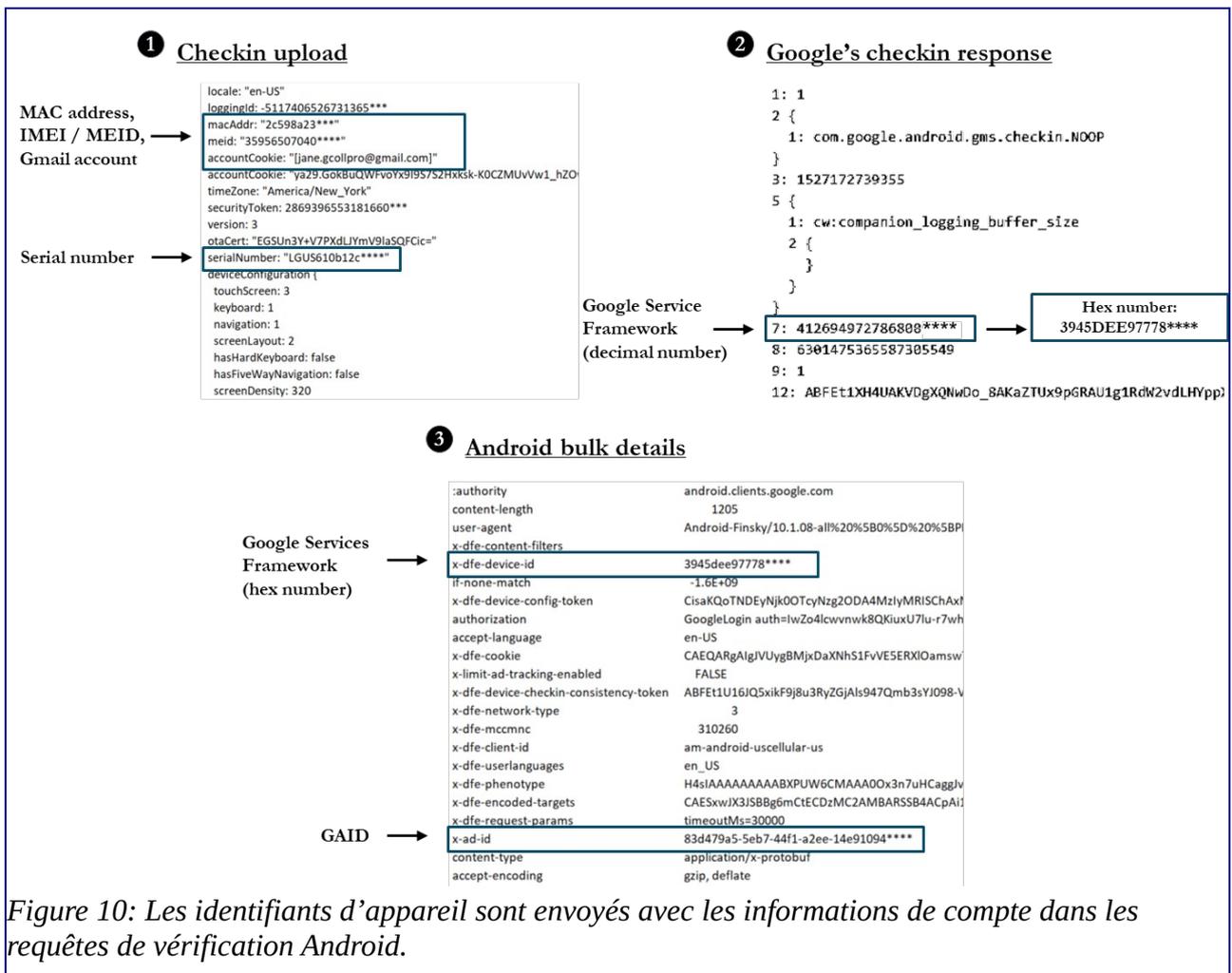
1) L'identificateur de publicité mobile peut être désanonymisé grâce aux données envoyées à Google par Android.

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 à la figure 10.

46 "Google privacy and terms", Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://policies.google.com/privacy/example/your-activity-on-other-sites-and-apps>

47 Google <https://myaccount.google.com/activitycontrols>



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)⁴⁸ qui est similaire à l'« Android ID »⁴⁹ (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.

48 "Difference between Android ID and device ID", Stack Exchange, décembre 2016, disponible à l'adresse <https://android.stackexchange.com/questions/162448/difference-between-android-id-and-device-id>

49 Patrick Ahlbrecht, "What's the difference between the GSF ID and the Android ID", Onyxbits, mars 2016, disponible à l'adresse <https://blog.onyxbits.de/whats-the-difference-between-the-gsf-id-and-the-android-id-208/>

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'utilisateur⁵⁰. 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 publicité⁵¹, car ces outils sont conçus principalement pour empêcher la présence de publicités pendant que les utilisateurs naviguent sur des pages web tierces⁵². La section suivante examine de plus près l'ampleur de cette collecte de données.

50 Une étape d'authentification supplémentaire permet au navigateur de l'utilisateur d'utiliser le même jeton d'authentification pour confirmer son identité sur d'autres services Google (à cause de ce processus, une identification à n'importe laquelle des applications Google active une identification automatique à toutes les autres dans la même session du navigateur).

51 "How many users block Google Analytics, measured in Google Analytics," Quantable, Dec. 2017, disponible à l'adresse <https://www.quantable.com/analytics/how-many-users-block-google-analytics/>

52 "Ad blocking : who blocks ads, why and how to win them back," iab., 2016, disponible à l'adresse <https://www.iab.com/wp-content/uploads/2016/07/IAB-Ad-Blocking-2016-Who-Blocks-Ads-Why-and-How-to-Win-Them-Back.pdf>

V. Quantité de données collectées lors d'une utilisation minimale des produits Google

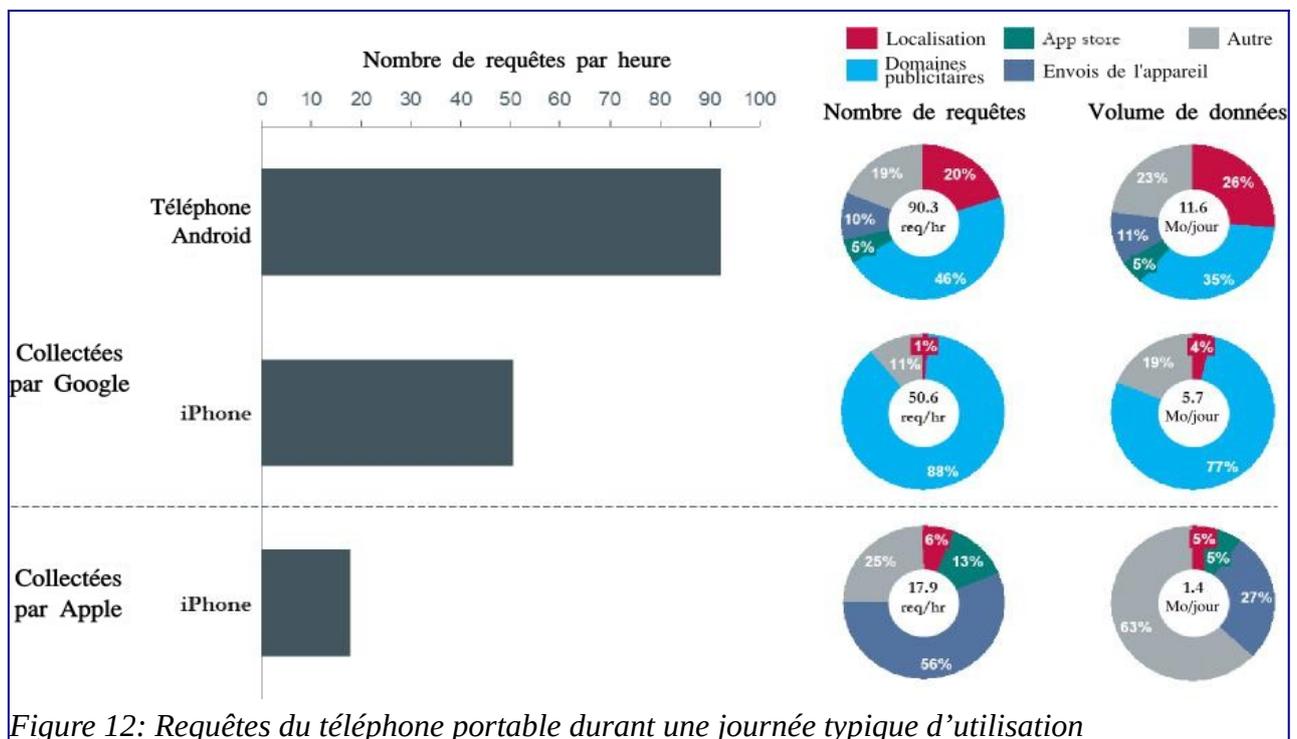
58. Cette section montre les détails de la collecte de données par Google à travers ses services de publication et d'annonces. Afin de comprendre une telle collecte de données, une expérience est réalisée impliquant un utilisateur qui se sert de son téléphone dans sa vie de tous les jours mais qui **évite** délibérément d'utiliser les produits Google (Search, Gmail, YouTube, Maps, etc.), exception faite du navigateur Chrome.

59. Pour que l'expérience soit aussi réaliste que possible, plusieurs études sur les usages de consommateurs^{53 et 54} ont été utilisées pour créer le profil d'usage journalier d'un utilisateur lambda. Ensuite, toutes les interactions directes avec les services Google ont été retirées du profil. La section IX.F dans les annexes liste les sites internet et applications utilisés pendant l'expérience.

60. L'expérience a été reproduite sur des appareils Android et iOS et les données HTTPS envoyées aux serveurs Google et Apple ont été tracées et analysées en utilisant une méthode similaire à celle expliquée dans la section précédente. Les résultats sont résumés dans la figure 12. Pendant la période de 24 h (qui inclut la période de repos nocturne), la majorité des appels depuis le téléphone Android ont été effectués vers les services Google de localisation et de publication de publicités (DoubleClick, Analytics). Google a enregistré la géolocalisation de l'utilisateur environ 450 fois, ce qui représente 1,4 fois le volume de l'expérience décrite dans la section III.C, qui se basait sur un téléphone immobile.

53 Une personne moyenne visitait 88 pages Web par jour en 2010. « Nielsen provides topline U.S. web data for March 2010 », Nielsen, avril 2010, disponible à l'adresse <http://www.nielsen.com/us/en/insights/news/2010/nielsen-provides-topline-u-s-web-data-for-march-2010.html>

54 55 % du trafic Web provient des appareils mobiles. Eric Enge, « Mobile vs desktop usage : mobile grows but desktop still a big player in 2017 », Stone Temple, avril 2017, disponible à l'adresse <https://www.stonetemple.com/mobile-vs-desktop-usage-mobile-grows-but-desktop-still-a-big-player-in-2017/>



61. Les serveurs de Google communiquent significativement moins souvent avec un appareil iPhone qu'avec Android (45 % moins). En revanche, le nombre d'appels aux régies publicitaires de Google reste les mêmes pour les deux appareils, c'est un résultat prévisible puisque l'utilisation de pages web et d'applications tierces était la même sur chacun des périphériques. À noter, une différence importante est que l'envoi de données de géolocalisation à Google depuis un appareil iOS est pratiquement inexistant. En absence des plateformes Android et Chrome – ou de l'usage d'un des autres produits de Google – Google perd significativement sa capacité à pister la position des utilisateurs.

62. Le nombre total d'appels aux serveurs Apple depuis un appareil iOS était bien moindre, seulement 19 % des appels aux serveurs de Google depuis l'appareil Android. De plus, il n'y a pas d'appels aux serveurs d'Apple liés à la publicité, ce qui pourrait provenir du fait que le modèle économique d'Apple ne dépend pas autant de la publicité que celui de Google. Même si Apple obtient bien certaines données de localisation des utilisateurs d'appareil iOS, le volume de données collectées est bien moindre (16 fois moins) que celui collecté par Google depuis Android.

63. Au total, les téléphones Android ont communiqué 11,6 Mo de données par jour (environ 350 Mo par mois) avec les serveurs de Google. En comparaison, l'iPhone n'a envoyé que la moitié de ce volume. La quantité de données spécifiques aux régies publicitaires de Google est restée pratiquement identique sur les deux appareils.

64. L'appareil iPhone a communiqué bien moins de données aux serveurs Apple que l'appareil Android n'a échangé avec les serveurs Google.

65. De manière générale, même en l'absence d'interaction utilisateur avec les applications Google les plus populaires, un utilisateur de téléphone Android muni du navigateur Chrome a tout de même tendance à envoyer une quantité non négligeable de données à Google, dont la majorité est liée à la

localisation et aux appels aux serveurs de publicité. Bien que, dans le cadre limité de cette expérience, un utilisateur d'iPhone soit protégé de la collecte des données de localisation par Google, Google recueille tout de même une quantité comparable de données liées à la publicité.

66. La section suivante décrit les données collectées par les applications les plus populaires de Google, telles que Gmail, Youtube, Maps et la recherche.

VI. Données collectées par les applications clés de Google destinées aux particuliers

67. Google a des dizaines de produits et services qui évoluent en permanence (une liste est disponible dans le tableau 4, section IX.B de l'annexe). On accède souvent à ces produits grâce à un compte Google (ou on l'y associe), ce qui permet à Google de relier directement les détails des activités de l'utilisateur de ses produits et services à un profil utilisateur. En plus des données d'usage de ses produits, Google collecte également des identifiants et des données de localisation liés aux appareils lorsqu'on accède aux services Google⁵⁵.

68. Certaines applications de Google (p.ex. YouTube, Search, Gmail et Maps) occupent une place centrale dans les tâches de base qu'une multitude d'utilisateurs effectuent quotidiennement sur leurs appareils fixes ou mobiles. Le tableau 2 décrit la portée de ces produits clés. Cette section explique comment chacune de ces applications majeures collecte les informations des utilisateurs.

Tableau 2 : Portée mondiale des principales applications Google

Produits	Utilisateurs actifs
Search	Plus d'un milliard d'utilisateurs actifs par mois, 90,6 % de part de marché des moteurs de recherche ⁵⁶
Youtube	Plus de 1,8 milliard d'utilisateurs inscrits et actifs par mois ⁵⁷
Maps	Plus d'un milliard d'utilisateurs actifs par mois ⁵⁸
Gmail	1,2 milliard d'utilisateurs enregistrés ⁵⁹

A. Recherche

69. Google Search est le moteur de recherche sur internet le plus populaire au monde⁶⁰, avec plus de 11 milliards de requêtes par mois aux États-Unis⁶¹. En plus de renvoyer un classement de pages web en réponse aux requêtes globales des utilisateurs, Google exploite d'autres outils basés sur la recherche, tels que Google Finance, Flights (vols), News (actualités), Scholar (recherche universitaire), Patents (brevets), Books (livres), Images, Videos et Hotels. Google utilise ses applications de recherche afin de collecter des données liées aux recherches, à l'historique de navigation ainsi qu'aux activités d'achats et de clics sur publicités. Par exemple, Google Finance collecte des informations sur le type d'actions que les utilisateurs peuvent suivre, tandis que Google Flight piste leurs réservations et recherches de voyage.

55 « Règles de confidentialité et conditions d'utilisation », consulté le 15 août 2018, <https://policies.google.com/privacy>

56 « Parts de marché des moteurs de recherche dans le monde » StatCounter Global Stats, avril 2018, <http://gs.statcounter.com/search-engine-market-share#monthly-201704-201804>

57 Devindra Hardawar, « YouTube accueille 1,8 milliards d'utilisateurs connectés chaque mois », Engadget, 3 mai 2018, <https://www.engadget.com/2018/05/03/youtube-1-8-billion-viewers/>

58 Rapport annuel Form 10 K de Google 10 K pour la SEC, 2017, https://abc.xyz/investor/pdf/20171231_alphabet_10K.pdf

59 Motek Moyen, « Gmail est très populaire mais Google persiste à ne pas corriger une faille de sécurité » Seeking Alpha, juillet 2017, <https://seekingalpha.com/article/4088241-gmail-popular-google-still-fix-security-vulnerability>

60 « Parts de marché des moteurs de recherche dans le monde », StatCounter Global Stats, avril 2018, <http://gs.statcounter.com/search-engine-market-share#monthly-201704-201804>

61 « Nombre de requêtes de recherche de bases réalisées par les moteurs de recherche aux États-Unis jusqu'en janvier 2018 (en milliards) », Statista, février 2018, <https://www.statista.com/statistics/265796/us-search-engines-ranked-by-number-of-core-searches/>

70. Dès lors que Search est utilisé, Google collecte les données de localisation par différents biais, sur ordinateur ou sur mobile, comme décrit dans les sections précédentes. Google enregistre toute l'activité de recherche d'un utilisateur ou utilisatrice et la relie à son compte Google si cette personne est connectée. La figure 13 montre un exemple d'informations collectées par Google sur une recherche utilisateur par mot-clé et la navigation associée.



Figure 13: Un exemple de collecte de données de recherche extrait de la page My Activity (Mon Activité) d'un utilisateur

71. Non seulement c'est le moteur de recherche par défaut sur Chrome et les appareils Google, mais Google Search est aussi l'option par défaut sur d'autres navigateurs internet et applications grâce à des arrangements de distribution. Ainsi, Google est récemment devenu le moteur de recherche par défaut sur le navigateur internet Mozilla Firefox⁶² dans des régions clés (dont les USA et le Canada), une position occupée auparavant par Yahoo. De même, Apple est passé de Microsoft Bing à Google pour les résultats de recherche via Siri sur les appareils iOS et Mac⁶³. Google a des accords similaires en place avec des OEM (fabricants d'équipement informatique ou électronique)⁶⁴, ce qui lui permet d'atteindre les consommateurs mobiles.

B. YouTube

72. YouTube met à disposition des utilisateurs et utilisatrices une plateforme pour la mise en ligne et la visualisation de contenu vidéo. Il attire plus de 180 millions de personnes rien qu'aux USA et a la particularité d'être le deuxième site le plus visité des États-Unis⁶⁵, juste derrière Google Search. Au sein des entreprises de streaming multimédia, YouTube possède près de 80 % de parts de marché en termes de visites mensuelles (comme décrit dans la figure 14). La quantité de contenu mis en ligne et

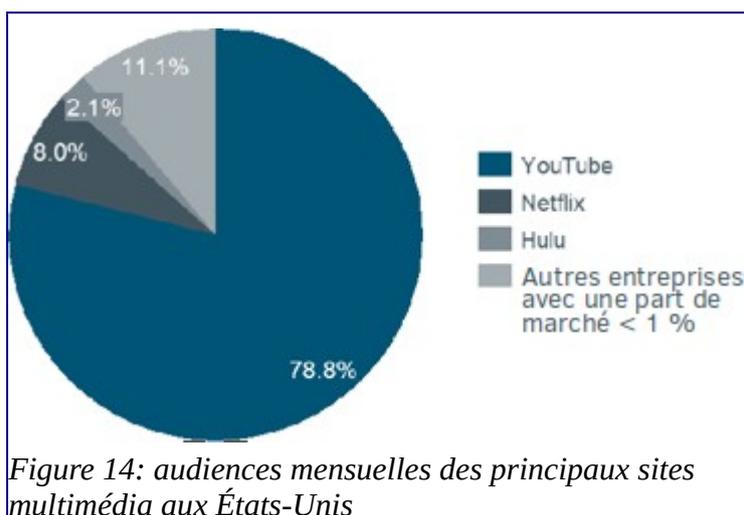
62 Denelle Dixon, « Firefox fournit le moteur de recherche Google par défaut aux États-Unis, au Canada, à Hong Kong et à Taiwan », The Mozilla Blog, 14 novembre 2017, <https://blog.mozilla.org/blog/2017/11/14/firefox-features-google-as-default-search-provider-in-the-u-s-canada-hong-kong-and-taiwan/>

63 Matthew Panzarino, « Apple passe de Bing à Google pour les résultats de recherches de Siri sur iOS et Spotlight sur Mac », 25 septembre 2017, <https://techcrunch.com/2017/09/25/apple-switches-from-bing-to-google-for-siri-web-search-results-on-ios-and-spotlight-on-mac/>

64 « Les accords de distribution des applications mobile Androids entre Google et les Fabricants d'équipement d'origine (OEM) ont fuités et révèlent beaucoup de restrictions strictes », Microsoft and Technology News, 13 février 2014, <https://mspoweruser.com/mobile-application-distribution-agreement/>

65 « Classement des sites aux États-Unis » Alexa, <https://www.alexa.com/topsites/countries/US>

visualisé sur YouTube est conséquente : 400 heures de vidéo sont mises en ligne chaque minute⁶⁶ et 1 milliard d'heures de vidéo sont visualisées quotidiennement sur la plateforme YouTube⁶⁷.



68

73. Les utilisateurs peuvent accéder à YouTube sur l'ordinateur (navigateur internet), sur leurs appareils mobiles (application et/ou navigateur internet) et sur Google Home (via un abonnement payant appelé YouTube Red). Google collecte et sauvegarde l'historique de recherche, l'historique de visualisation, les listes de lecture, les abonnements et les commentaires aux vidéos. La date et l'horaire de chaque activité sont ajoutés à ces informations.

74. Si un utilisateur se connecte à son compte Google pour accéder à n'importe quelle application Google via un navigateur internet (par ex. Chrome, Firefox, Safari), Google reconnaît l'identité de l'utilisateur, même si l'accès à la vidéo est réalisé par un site hors Google (ex. : vidéos YouTube lues sur cnn.com). Cette fonctionnalité permet à Google de pister l'utilisation YouTube d'un utilisateur à travers différentes plateformes tierces. 15 montre un exemple de données YouTube collectées.

66 « Nombres d'heures de vidéos postées sur YouTube chaque minute en date de juillet 2015, » Statista, July 2015, <https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/>

67 Darrell Etherington, « Les utilisateurs regardent maintenant 1 milliard d'heures de vidéos sur YouTube par jour », TeachCrunch, 28 février 2017, <https://techcrunch.com/2017/02/28/people-now-watch-1-billion-hours-of-youtube-per-day/>

68 « Sites multimédia dominants aux États-Unis en novembre 2016, en fonction des parts de marchés de visites », Statista, décembre 2016, <https://www.statista.com/statistics/266201/us-market-share-of-leading-internet-video-portals/>

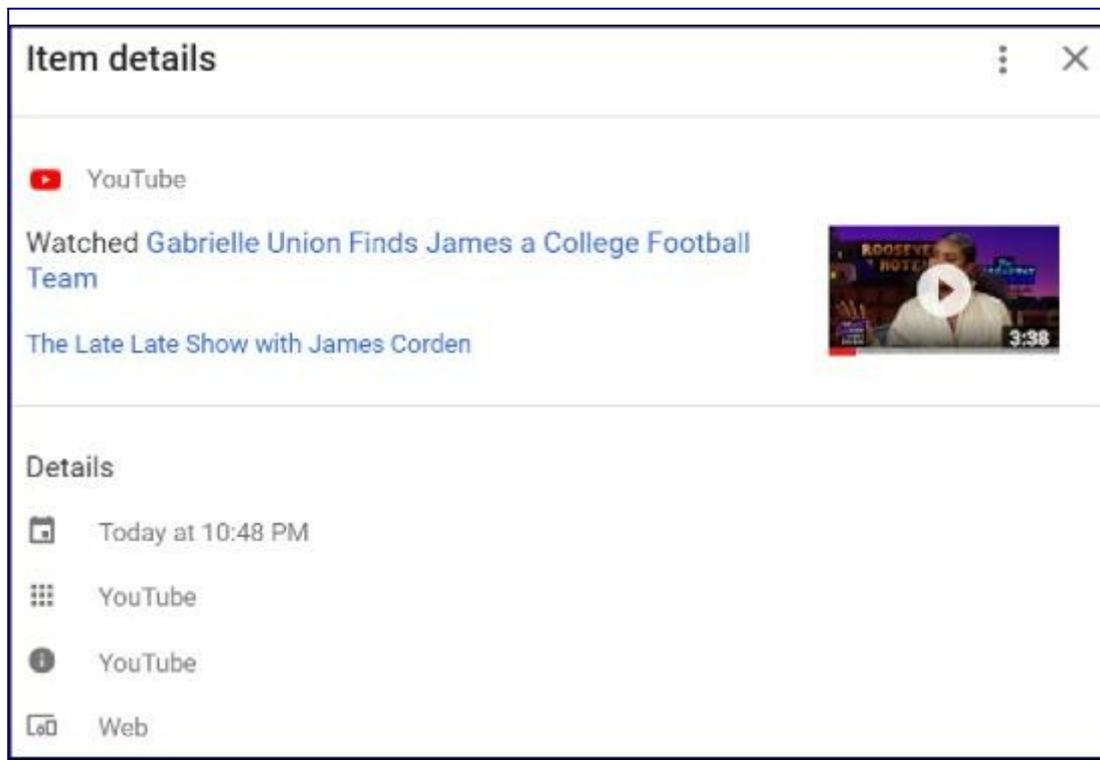


Figure 15: Exemple de collecte de données YouTube dans My Activity (Mon Activité)

75. Google propose également un produit YouTube différencié pour les enfants, appelé YouTube Kids, dans l'intention d'offrir une version « familiale » de YouTube avec des fonctionnalités de contrôle parental et de filtres vidéos. Google collecte des informations de YouTube Kids, notamment le type d'appareil, le système d'exploitation, l'identifiant unique de l'appareil, les informations de journalisation et les détails d'utilisation du service. Google utilise ensuite ces informations pour fournir des annonces publicitaires limitées, qui ne sont pas cliquables et dont le format, la durée et le site sont limités⁶⁹.

C. Maps

76. Maps est l'application phare de navigation routière de Google. Google Maps peut déterminer les trajets et la vitesse d'un utilisateur et ses lieux de fréquentation régulière (ex. : domicile, travail, restaurants et magasins). Cette information donne à Google une idée des intérêts (ex. : préférences d'alimentation et d'achats), des déplacements et du comportement de l'utilisateur.

77. Maps utilise l'adresse IP, le GPS, le signal cellulaire et les points d'accès au Wi-Fi pour calculer la localisation d'un appareil. Les deux dernières informations sont collectées par le biais de l'appareil où Maps est utilisé, puis envoyées à Google pour évaluer la localisation via son interface de localisation (Location API). Cette interface fournit de nombreux détails sur un utilisateur, dont les coordonnées géographiques, son état stationnaire ou en mouvement, sa vitesse et la détermination probabiliste de son mode de transport (ex. : en vélo, voiture, train, etc.).

78. Maps sauvegarde un historique des lieux qu'un utilisateur connecté à Maps par son compte Google a visités. La figure 16. montre un exemple d'un tel historique⁷⁰. Les points rouges indiquent

69 « Faire de la publicité sur l'application YouTube Kids », Google, consulté le 15 août 2018, <https://support.google.com/youtube/answer/6168681?hl=en>

70 « My Activity », Google, <https://myactivity.google.com/myactivity>

les coordonnées géographiques recueillies par Maps lorsque l'utilisateur se déplace ; les lignes bleues représentent les projections de Maps sur le trajet réel de l'utilisateur.



79. La précision des informations de localisation recueillies par les applications de navigation routière permet à Google de non seulement cibler des audiences publicitaires, mais l'aide aussi à fournir des annonces publicitaires aux utilisateurs lorsqu'ils s'approchent d'un magasin⁷¹. Google Maps utilise de plus ces informations pour générer des données de trafic routier en temps réel⁷².

D. Gmail

80. Gmail sauvegarde tous les messages (envoyés et reçus), le nom de l'expéditeur, son adresse email et la date et l'heure des messages envoyés ou reçus. Puisque Gmail représente pour beaucoup un répertoire central pour la messagerie électronique, il peut déterminer leurs intérêts en scannant le contenu de leurs courriels, identifier les adresses de commerçants grâce à leurs courriels publicitaires ou les factures envoyées par message électronique, et connaître l'agenda d'un utilisateur (ex. : réservations à dîner, rendez-vous médicaux...). Étant donné que les utilisateurs utilisent leur identifiant Gmail pour des plateformes tierces (Facebook, LinkedIn...), Google peut analyser tout contenu qui leur parvient sous forme de courriel (ex. : notifications, messages).

81. Depuis son lancement en 2004 jusqu'à la fin de l'année 2017 (au moins), Google peut avoir analysé le contenu des courriels Gmail pour améliorer le ciblage publicitaire et les résultats de recherche ainsi que ses filtres de pourriel. Lors de l'été 2016, Google a franchi une nouvelle étape et a modifié sa politique de confidentialité pour s'autoriser à fusionner les données de navigation, autrefois anonymes, de sa filiale DoubleClick (qui fournit des publicités personnalisées sur internet) avec les données d'identification personnelles qu'il amasse à travers ses autres produits, dont

71 « The Home Depot earns 8X in-store ROI with mobile display ads », Google, septembre 2016, <https://www.thinkwithgoogle.com/intl/en-aunz/advertising-channels/mobile/home-depot-roi-mobile-display-ads/>

72 « Google Map's real-time traffic layer », Spatial Unlimited, mars 2011, <https://shreerangpatwardhan.blogspot.com/2011/03/google-maps-real-time-traffic-layer.html>

Gmail⁷³. Le résultat : « les annonces publicitaires DoubleClick qui pistent les gens sur Internet peuvent maintenant leur être adaptées sur mesure, en se fondant sur les mots-clés qu'ils ont utilisés dans leur messagerie Gmail. Cela signifie également que Google peut à présent reconstruire le portrait complet d'une utilisatrice ou utilisateur par son nom, en fonction de tout ce qui est écrit dans ses courriels, sur tous les sites visités et sur toutes les recherches menées. »⁷⁴

82. Vers la fin de l'année 2017, Google a annoncé qu'il arrêterait la personnalisation des publicités basées sur les messages Gmail⁷⁵. Cependant, Google a annoncé récemment qu'il continue à analyser les messages Gmail pour certaines raisons⁷⁶.

73 Julia Angwin, « Google has quietly dropped ban on personally identifiable web tracking, » ProPublica, 21 octobre 2016, <https://www.propublica.org/article/google-has-quietly-dropped-ban-on-personally-identifiable-web-tracking>

74 Suzanne Monyak, « Google a modifié un élément majeur de sa politique de confidentialité il y a quatre mois et personne ne s'en est rendu compte », Slate, 21 octobre 2016, http://www.slate.com/blogs/future_tense/2016/10/21/google_changed_a_major_privacy_policy_and_no_one_really_noticed.html

75 Mark Bergen, « Google va arrêter de lire vos courriels pour les publicités Gmail », Bloomberg, June 23, 2017, <https://www.bloomberg.com/news/articles/2017-06-23/google-will-stop-reading-your-emails-for-gmail-ads>

76 Ben Popken, « Google vend le futur, alimenté par vos données personnelles », NBC News, 10 mai 2018, <https://www.nbcnews.com/tech/tech-news/google-sells-future-powered-your-personal-data-n870501>

VII. Des produits avec un haut potentiel futur d'agrégation de données

83. Google a d'autres produits qui pourraient être adoptés par le marché et pourraient bientôt servir à la collecte de données, tels que AMP, Photos, Chromebook Assistant et Google Pay. Il faut ajouter à cela que Google est capable d'utiliser les données provenant de partenaires pour collecter les informations de l'utilisateur. La section suivante les décrit plus en détail.

84. Il existe également d'autres applications Google qui peuvent ne pas être largement utilisées. Toutefois, par souci d'exhaustivité, la collecte de données par leur intermédiaire est présentée dans la section VIII.B de l'annexe.

A. Pages optimisées pour les mobiles (AMP)

85. Les Pages optimisées pour les mobiles (AMP) sont une initiative *open source* menée par Google pour réduire le temps de chargement des sites Web et des publicités. AMP convertit le HTML standard et le code JavaScript en une version simplifiée développée par Google⁷⁷ qui stocke les pages validées dans un cache des serveurs du réseau Google pour un accès plus rapide⁷⁸. AMP fournit des liens vers les pages grâce aux résultats de la recherche Google mais également via des applications tierces telles que LinkedIn et Twitter. D'après AMP : « L'écosystème AMP compte 25 millions de domaines, plus de 100 fournisseurs de technologie et plateformes de pointe qui couvrent les secteurs de la publication de contenu, les publicités, le commerce en ligne, les petits commerces, le commerce local, etc. »⁷⁹

86. La figure 17a décrit les étapes menant à la fourniture d'une page AMP accessible via la recherche Google. Il convient de noter que le fournisseur de contenu à travers AMP n'a pas besoin de fournir ses propres caches serveur, car c'est quelque chose que Google fournit pour garantir un délai optimal de livraison aux utilisateurs. Dans la mesure où le cache AMP est hébergé sur les serveurs de Google, lors d'un clic sur un lien AMP produit par la recherche Google, le nom de domaine vient du domaine google.com et non pas du domaine du fournisseur. Ceci est montré grâce aux captures prises lors d'un exemple de recherche de mots clés dans la figure 17b.

77 « AMP HTML specification » disponible à l'adresse <https://www.ampproject.org/docs/fundamentals/spec>

78 « Charger des pages AMP rapidement avec le Cache AMP Google », visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://developers.google.com/amp/cache>

79 Un effort open source pour améliorer, pour tous, l'écosystème du contenu AMP, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://www.ampproject.org/learn/overview>

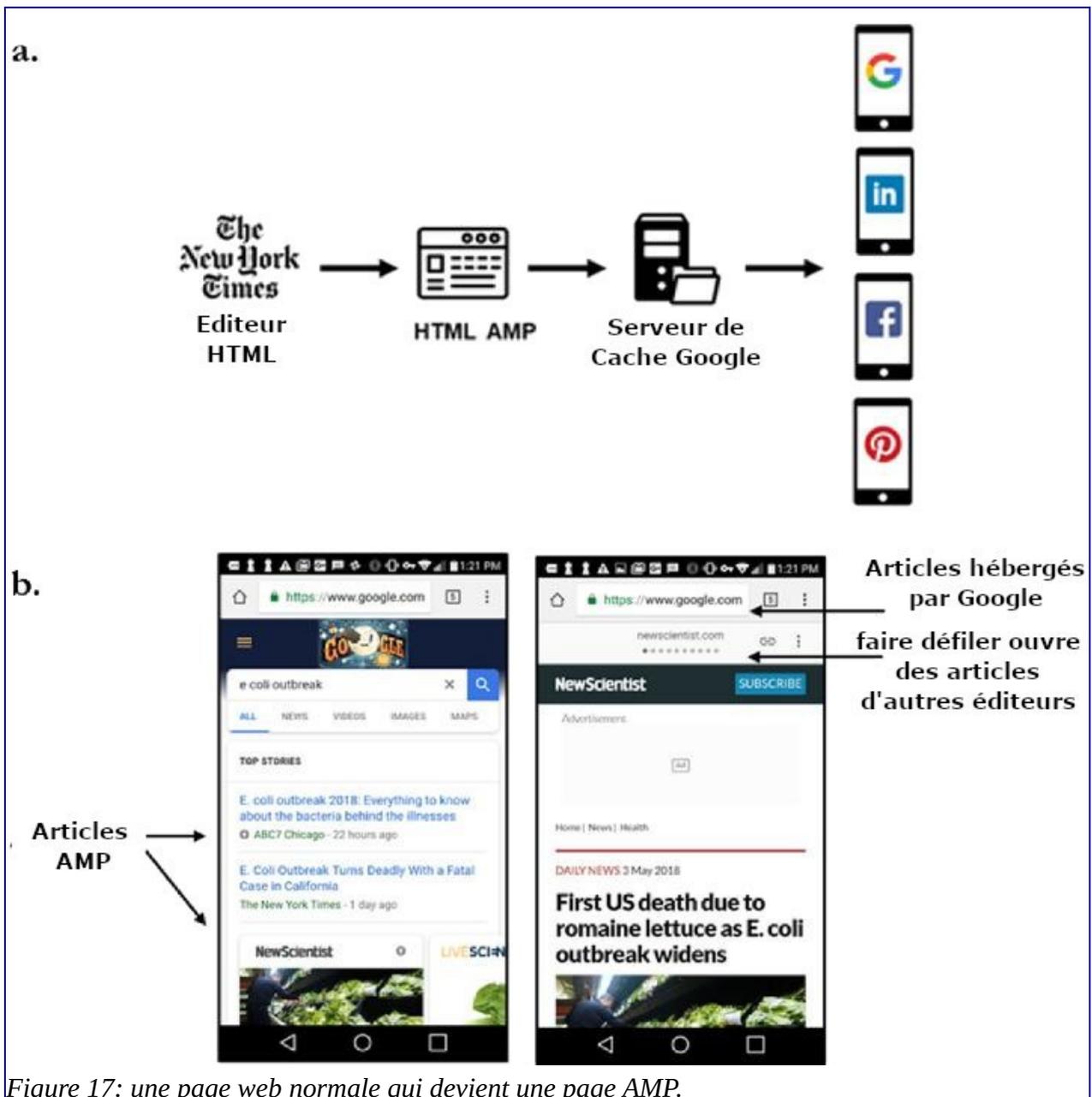


Figure 17: une page web normale qui devient une page AMP.

87. Les utilisateurs peuvent accéder au contenu depuis de multiples fournisseurs dont les articles apparaissent dans les résultats de recherche pendant qu'ils naviguent dans le carrousel AMP, tout en restant dans le domaine de Google. En effet, le cache AMP opère comme un réseau de distribution de contenu (RDC, ou CDN en anglais) appartenant à Google et géré par Google.

88. En créant un outil *open source*, complété avec un CDN, Google a attiré une large base d'utilisateurs à qui diffuser les sites mobiles et la publicité et cela constitue une quantité d'information significative (par ex. : le contenu lui-même, les pages vues, les publicités, et les informations de celui à qui ce contenu est fourni). Toutes ces informations sont disponibles pour Google parce qu'elles sont collectées sur les serveurs CDN de Google, fournissant ainsi à Google beaucoup plus de données que par tout autre moyen d'accès.

89. L'AMP est très centré sur l'utilisateur, c'est-à-dire qu'il offre aux utilisateurs une expérience de navigation beaucoup plus rapide et améliorée sans l'encombrement des fenêtres pop-up et des barres

latérales. Bien que l'AMP représente un changement majeur dans la façon dont le contenu est mis en cache et transmis aux utilisateurs, la politique de confidentialité de Google associée à l'AMP est assez générale⁸⁰. En particulier, Google est en mesure de recueillir des informations sur l'utilisation des pages Web (par exemple, les journaux de serveur et l'adresse IP) à partir des requêtes envoyées aux serveurs de cache AMP. De plus, les pages standards sont converties en AMP via l'utilisation des API AMP⁸¹. Google peut donc accéder à des applications ou à des sites Web (« clients API ») et utiliser toute information soumise par le biais de l'API conformément à ses politiques générales⁸².

90. Comme les pages Web ordinaires, les pages Web AMP pistent les données d'utilisation via Google Analytics et DoubleClick. En particulier, elles recueillent des informations sur les données de page (par exemple : domaine, chemin et titre de page), les données d'utilisateur (par exemple : ID client, fuseau horaire), les données de navigation (par exemple : ID et référence de page uniques), l'information du navigateur et les données sur les interactions et les événements⁸³. Bien que les modes de collecte de données de Google n'aient pas changé avec l'AMP, la quantité de données recueillies a augmenté puisque les visiteurs passent 35 % plus de temps sur le contenu Web qui se charge avec Google AMP que sur les pages mobiles standard⁸⁴.

B. Google Assistant

91. Google Assistant est un assistant personnel virtuel auquel on accède par le biais de téléphones mobiles et d'appareils dits intelligents. C'est un assistant virtuel populaire, comme Siri d'Apple, Alexa d'Amazon et Cortana de Microsoft⁸⁵. Google Assistant est accessible via le bouton d'accueil des appareils mobiles sous Android 6.0 ou versions ultérieures. Il est également accessible via une application dédiée sur les appareils iOS⁸⁶, ainsi que par l'intermédiaire de haut-parleurs intelligents, tel Google Home, qui offre de nombreuses fonctions telles que l'envoi de messages, la recherche de courriels, le contrôle de la musique, la recherche de photos, les réponses aux questions sur la météo ou la circulation, et le contrôle des appareils domestiques intelligents⁸⁷.

92. Google collecte toutes les requêtes de Google Assistant, qu'elles soient audio ou saisies au clavier. Il collecte également l'emplacement où la requête a été effectuée. La figure 18 montre le contenu d'une requête enregistrée par Google. Outre son utilisation via les haut-parleurs de Google Home, Google Assistant est activé sur divers autres haut-parleurs produits par des tiers (par exemple, les casques sans fil de Bose). Au total, Google Assistant est disponible à l'adresse plus de 400

80 « AMP dans les Règles de confidentialité et conditions d'utilisation de Google », visité la dernière fois le 15 août 2018, disponible à l'adresse <https://developers.google.com/amp/cache/policies>

81 « Lien vers du contenu AMP », Google, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://developers.google.com/amp/cache/use-amp-url>

82 « Règles de confidentialité et conditions d'utilisation des APIs Google », visité la dernière fois le 15 août 2018, disponible à l'adresse <https://developers.google.com/terms>

83 « suivi analytique des pages mobiles accélérées (AMP) » Google, visité la dernière fois le 15 août 2018, disponible à l'adresse https://support.google.com/analytics/answer/6343176?hl=en&ref_topic=7378717

84 John Saroff, « Nouvelle norme de vitesse pour l'engagement utilisateur » Chartbeat, 5 juin 2017, disponible à l'adresse <http://blog.chartbeat.com/2017/06/05/the-new-speed-of-mobile-engagement>

85 Tripp Mickle, « Je ne suis pas sûr de comprendre' – Comment Siri de Apple à perdu son Mojo » The Wall Street Journal, 7 juin 2017, disponible à l'adresse <https://www.wsj.com/articles/apples-siri-once-an-original-now-struggles-to-be-heard-above-the-crowd-1496849095>

86 « Google Assistant, » Google, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://assistant.google.com/platforms/phones/>

87 « Google Assistant, » Google, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://assistant.google.com/platforms/phones/>

millions d'appareils⁸⁸. Google peut collecter des données via l'ensemble de ces appareils puisque les requêtes de l'Assistant passent par les serveurs de Google.

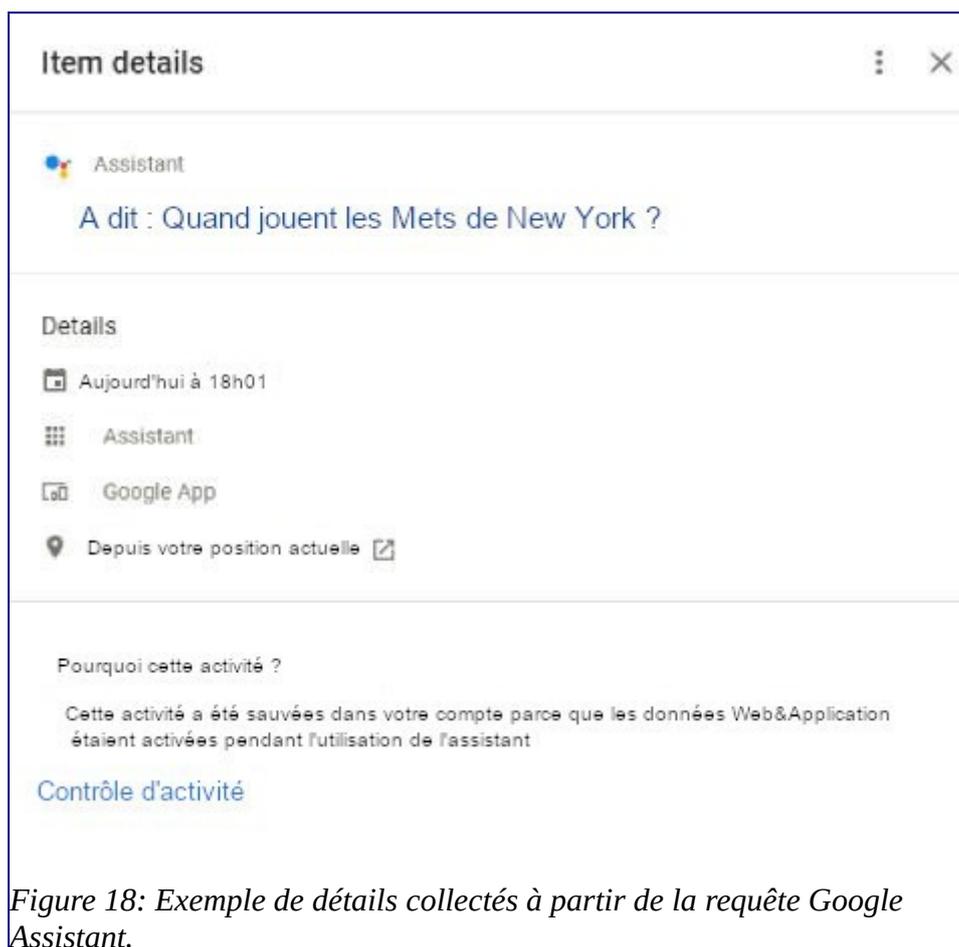


Figure 18: Exemple de détails collectés à partir de la requête Google Assistant.

C. Photos

93. Google Photos est utilisé par plus de 500 millions de personnes dans le monde et stocke plus de 1,2 milliard de photos et vidéos chaque jour⁸⁹. Google enregistre l'heure et les coordonnées GPS de chaque photo prise. Google télécharge des images dans le Google cloud et effectue une analyse d'images pour identifier un large éventail d'objets, tels que les modes de transport, les animaux, les logos, les points de repère, le texte et les visages⁹⁰. Les capacités de détection des visages de Google permettent même de détecter les états émotionnels associés aux visages dans les photos téléchargées et stockées dans leur cloud⁹¹.

88 Scott Huffman, « de nouveaux appareils et plus encore : Qu'est qui est en stock cette année pour l'assistant Google », Google, 9 janvier 2018, disponible à l'adresse <https://www.blog.google/products/assistant/new-devices-more-google-assistant-ces-2018>

89 Anil Sabharwal, « 500 millions de personnes utilisent Google Photos et 3 nouvelles façons de partager », Google, visité la dernière fois le 9 janvier 2018, disponible à l'adresse <https://blog.google/products/photos/google-photos-500-million-new-sharing/>

90 "API Cloud vision," Google, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://cloud.google.com/vision/>

91 "API Cloud vision," Google, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://cloud.google.com/vision/>

94. Google Photos effectue cette analyse d'image par défaut lors de l'utilisation du produit, mais ne fera pas de distinction entre les personnes, sauf si l'utilisateur donne l'autorisation à l'application⁹². Si un utilisateur autorise Google à regrouper des visages similaires, Google identifie différentes personnes à l'aide de la technologie de reconnaissance faciale et permet aux utilisateurs de partager des photos grâce à sa technologie de « regroupement de visages »^{93 et 94}. Des exemples des capacités de classification d'images de Google avec et sans autorisation de regroupement des visages de l'utilisateur sont présentés dans la figure 19. Google utilise Photos pour assembler un vaste ensemble d'informations d'identifications faciales, qui a récemment fait l'objet de poursuites judiciaires⁹⁵ de la part de certains États.



Figure 19: Exemple de reconnaissance d'images dans Google Photos.

D. Chromebook

95. Chromebook est la tablette-ordinateur de Google qui fonctionne avec le système d'exploitation Chrome (Chrome OS) et permet aux utilisateurs d'accéder aux applications sur le *cloud*. Bien que Chromebook ne détienne qu'une très faible part du marché des PC, il connaît une croissance rapide, en particulier dans le domaine des appareils informatiques pour la catégorie K-12, où il détenait 59,8 % du marché au deuxième trimestre 2017⁹⁶. La collecte de données de Chromebook est similaire à celle du navigateur Google Chrome, qui est décrite dans la section II.A. Chromebooks

92 « Trouver des personnes, des choses et des lieux dans vos photos », API Cloud vision, Google, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://support.google.com/photos/answer/6128838?co=GENIE.Platform%3DAndroid&hl=en>

93 Anil Sabharwal, « 500 millions de personnes utilisent Google Photos et 3 nouvelles façons de partager », Google, visité la dernière fois le 9 janvier 2018, disponible à l'adresse <https://blog.google/products/photos/google-photos-500-million-new-sharing/>

94 « partagez votre bibliothèque de photos avec un partenaire » Google, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://support.google.com/photos/answer/7378858#filterbyface>

95 Amy Korte, « La Cour fédérale de l'Illinois a statué que la poursuite en matière de protection de la vie privée biométrique contre Google peut aller de l'avant », Illinois Privacy,

96 « Les ventes d'ordinateur portables aux établissements pré universitaires commencent à baisser au fur et à mesure que le marché entre dans le cycle de remplacement » Future Source Consulting, 6, décembre 2017, disponible à l'adresse <https://www.futuresource-consulting.com/Press-Q3-2017-Mobile-PC-Sales-in-Éducation-1217.html>

permet également aux cookies de Google et de domaines tiers de pister l'activité de l'utilisateur, comme pour tout autre ordinateur portable ou PC.

96. De nombreuses écoles de la maternelle à la terminale utilisent des Chromebooks pour accéder aux produits Google via son service GSuite for Éducation. Google déclare que les données recueillies dans le cadre d'une telle utilisation ne sont pas utilisées à des fins de publicité ciblée⁹⁷. Toutefois, les étudiants reçoivent des publicités s'ils utilisent des services supplémentaires (tels que YouTube ou Blogger) sur les Chromebooks fournis par leur établissement d'enseignement.

E. Google Pay

97. Google Pay est un service de paiement numérique qui permet aux utilisateurs de stocker des informations de carte de crédit, de compte bancaire et de PayPal pour effectuer des paiements en magasin, sur des sites Web ou dans des applications utilisant Google Chrome ou un appareil Android connecté⁹⁸. Pay est le moyen par lequel Google collecte les adresses et numéros de téléphone vérifiés des utilisateurs, car ils sont associés aux comptes de facturation. En plus des renseignements personnels, Pay recueille aussi des renseignements sur la transaction, comme la date et le montant de la transaction, l'emplacement et la description du marchand, le type de paiement utilisé, la description des articles achetés, toute photo qu'un utilisateur choisit d'associer à la transaction, les noms et adresses électroniques du vendeur et de l'acheteur, la description du motif de la transaction par l'utilisateur et toute offre associée à la transaction⁹⁹. Google traite ses informations comme des informations personnelles en fonction de sa politique générale de confidentialité. Il peut donc utiliser ces informations sur tous ses produits et services pour fournir de la publicité très ciblée¹⁰⁰. Les paramètres de confidentialité de Google l'autorisent par défaut à utiliser ces données collectées¹⁰¹.

F. Données d'utilisateurs collectées auprès de fournisseurs de données tiers

98. Google collecte des données de tiers en plus des informations collectées directement à partir de leurs services et applications. Par exemple, en 2014, Google a annoncé qu'il commencerait à suivre les ventes dans les commerces réels en achetant des données sur les transactions par carte bancaire. Ces données couvraient 70 % de toutes les opérations de crédit et de débit aux États-Unis¹⁰². Elles contenaient le nom de l'individu, ainsi que l'heure, le lieu et le montant de son achat¹⁰³.

99. Les données de tiers sont également utilisées pour aider Google Pay, y compris les services de vérification, les informations résultant des transactions Google Pay chez les commerçants, les

97 « Chromebooks vie privée et sécurité », Google, visité la dernière fois le 15 août 2018, disponible à l'adresse https://drive.google.com/file/d/0B_OTXR_u3RbcFB3Y01xUVhaalU/view

98 "Google Pay," Google, visité la dernière fois le 15 août 2018, disponible à <https://pay.google.com/about/>

99 « Avis de confidentialité de Google Payments » Google, 23 septembre 2018, disponible à l'adresse https://payments.google.com/payments/apis-secure/u/0/get_legal_document?ldo=0&ldt=privacynotice&ldl=fr

100 « Avis de confidentialité de Google Payments » Google, 23 septembre 2018, disponible à l'adresse https://payments.google.com/payments/apis-secure/u/0/get_legal_document?ldo=0&ldt=privacynotice&ldl=fr

101 "Google payments center," Google, disponible à <https://payments.google.com/payments/home?page=privacySettings#privacySettings>

102 « Google envisage de tracer les paiement via carte bleue » BBC, 26 mai 2017, disponible à l'adresse <http://www.bbc.com/news/technology-40027706>

103 Elizabeth Dwoskin and Craig Timberg, « Google sait maintenant quand ses utilisateurs vont en magasins et réalisent un achat » The Washington Post 23 mai 2017, disponible à l'adresse https://www.washingtonpost.com/news/the-switch/wp/2017/05/23/google-now-knows-when-you-are-at-a-cash-register-and-how-much-you-are-spending/?utm_term=.281715f8e215

méthodes de paiement, l'identité des émetteurs de cartes, les informations concernant l'accès aux soldes du compte de paiement Google, les informations de facturation des opérateurs et transporteurs et les rapports des consommateurs¹⁰⁴. Pour les vendeurs, Google peut obtenir des informations des organismes de crédit aux particuliers ou aux entreprises.

100. Bien que les informations des utilisateurs tiers que Google reçoit actuellement soient de limitées, cela a déjà attiré l'attention des autorités gouvernementales. Par exemple, la FTC a annoncé une injonction contre Google en juillet 2017 concernant la façon dont la collecte par Google de données sur les achats des consommateurs porte atteinte à la vie privée électronique¹⁰⁵. L'injonction conteste l'affirmation de Google selon laquelle il peut protéger la vie privée des consommateurs tout au long du processus en utilisant son algorithme. Bien que d'autres mesures n'aient pas encore été prises, l'injonction de la FTC est un exemple des préoccupations du public quant à la quantité de données que Google recueille sur les consommateurs.

VIII. CONCLUSION

101. Google compte un pourcentage important de la population mondiale parmi ses clients directs, avec de multiples produits en tête de leurs marchés mondiaux et de nombreux produits qui dépassent le milliard d'utilisateurs actifs par mois. Ces produits sont en mesure de recueillir des données sur les utilisateurs au moyen d'une variété de techniques qui peuvent être difficiles à comprendre pour un utilisateur moyen. Une grande partie de la collecte de données de Google a lieu lorsque l'utilisateur n'utilise *aucun* de ses produits directement. L'ampleur d'une telle collecte est considérable, en particulier sur les appareils mobiles Android. Et bien que ces informations soient généralement recueillies sans identifier un utilisateur unique, Google a la possibilité d'utiliser les données recueillies auprès d'autres sources pour *désanonymiser* une telle collecte.

104 « Avis de confidentialité de Google Payments » Google, 23 septembre 2018, disponible à l'adresse

https://payments.google.com/payments/apis-secure/u/0/get_legal_document?ldo=0&ldt=privacynotice&ldl=fr

105 « Plainte de la FTC, demande d'enquête, injonction et autres mesures de redressement présentées par le Electronic Privacy Information Center », disponible à l'adresse <https://epic.org/privacy/ftc/google/EPIC-FTC-Google-Purchase-Tracking-Complaint.pdf>

IX. ANNEXE

A. Caractérisation des collections de données actives vs passives à partir d'une journée-type d'un utilisateur

index	Description	Collecte Active	Collecte passive
1	Se prépare le matin en écoutant Google Play Music	préférences musicales	Lieu où se trouve la personne le matin
2	Dépose les enfants à la garderie avant de se rendre au travail.		S'est rendu à pied à une garderie
3	Lit les informations pendant qu'il se rend au travail dans le métro		Se déplace en métro Pages d'informations visitées
4	Recherche de médicaments contre le rhume dans le métro	Enregistre les requêtes de recherche	Se déplace en métro
5	Marche de la station de métro à son travail		trajet métro-travail
6	Utilise Maps pour trouver un nouveau lieu pour déjeuner	Destination entrée dans Maps	préférences alimentaires
7	Achète du café chez Starbucks à l'aide de son application Starbucks.		Marche jusqu'à un Starbucks Ouvre l'application Starbucks
8	Planifie un rendez-vous chez le médecin, Google crée un événement Calendrier à partir du courriel de confirmation.		Détails du de l'événement rendez-vous chez le médecin
9	Marche jusqu'à Walgreens (chaîne de pharmacie aux États-Unis [NDT]) et achète des médicaments contre le rhume en utilisant Google Pay	Détails des achats	Marche jusqu'à Walgreens
10	Prend un Uber de son travail pour revenir à la maison		Trajet jusqu'à la maison en utilisant une voiture Utilisation de l'application Uber
11	Regarde les hôtels sur Expedia pour partir un		Interaction avec des pages web via les cookies DoubleClick et Google Analytics

	week-end		
12	Utilise de Google Home pour jouer de la musique à ses enfants	Recherche Google Home	Localisation de Google Home
13	Regarde des vidéos sur Youtube	Activité Youtube	

Tableau 3 : collecte de données Google active et passive

B. Liste des produits Google

Catégorie	Produits
1. Applications	<u>Regarder, écouter et jouer</u> Youtube, Google Play musique, Chromecast, Google Play films et télévision
	<u>Navigateur</u> Chrome
	<u>Recherche</u> Recherche, Finance, Vols, Actualités, Recherche universitaire, Brevets, Livres, Images, Vidéos, Hôtels
	<u>Navigation</u> Maps, Waze
	<u>Outils de productivité</u> Drive, Docs, Sheets, Slides, Forms
	<u>Social & communications</u> Gmail, Allo, Hangouts, Duo, Google+, Translate
	<u>Stockage et organisation</u> Photos, Contacts, Calendar, Keep
	<u>Assistant personnel</u> Google voice assistant
2. Systèmes d'exploitation	<u>Android</u> Phones, Wear, Auto
	<u>Chrome</u> Chromebook
3. Services	Fiber (Service non disponible en France [NDT]), DNS, Projet Fi, Google Pay
4. Appareils	Home, Wi-Fi router, Chromecast, Nest, Daydream View

Tableau 4 : Liste des produits Google

C. Collecte de données via d'autres produits majeurs de Google

a) Google Play Music, Play Movies et TV

102. Google Play musique, Play films et la télévision à la demande sont des services à la demande qui proposent du streaming de musique, des podcasts, des séries télé, et des films. Ces services peuvent être vus comme les équivalents Google de Apple iTunes. Comme YouTube, ces services collectent des informations sur la recherche, l'achat / la location / la lecture de contenu, des informations sur la localisation géographique de l'utilisateur (grâce à son adresse IP), et des informations sur l'appareil.

b) Waze

103. Waze a été acquis par Google en 2013, et fonctionne comme une filiale de Google. Contrairement à Maps, Waze est une application participative dans laquelle les données des utilisateurs (comme les coordonnées GPS, le temps de trajet, les informations sur le trafic, les accidents, les contrôles routiers, les routes fermées et les travaux) sont analysées pour fournir en temps réel des mises à jour du de l'itinéraire et des conditions de voyage. Outre les informations de localisation collectées via le terminal mobile, Waze collecte des informations sur l'utilisation de ses services à partir du terminal sur lequel il est installé, incluant le nom du terminal, le système d'exploitation, les pages web visitées, les informations vues dans l'application, le contenu de l'application utilisé / créé, et les publicités vues et cliquées.

104. Waze fonctionne comme un réseau social et fournit aux utilisateurs la possibilité de devenir ami avec d'autres utilisateurs, et de créer des communautés en ligne de conducteurs locaux¹⁰⁶. Les utilisateurs peuvent lier leurs contacts téléphoniques ou leurs comptes Facebook ou Twitter, que Waze utilise ensuite pour établir des liens avec les amis qui utilisent aussi le service de Waze. Dans l'ensemble Waze fournit à Google la possibilité d'accéder à des données utilisateurs en temps réel, et également à des informations sur les liens qui peuvent ou non être utilisés dans un compte Google.

c) Google Docs et Drive

105. Les outils de productivité de Google (Docs, Sheets, Slides, Forms, and Drive, qui font partie de la G suite ensemble plus large de produits) sont des applications cloud utilisées par les particuliers et les entreprises. Les règles de confidentialité de G Suite¹⁰⁷ de Google interdisent à Google de scanner les données stockées à des fins publicitaires sur les versions d'entreprise. En revanche, les versions gratuites de ces outils sont régies par la politique générale de confidentialité de Google, de sorte que Google peut accéder aux informations pour le ciblage publicitaire.

106. Après l'inscription d'un individu à un compte Google, Google fournit un espace de stockage gratuit (actuellement 15 Go) sur le disque dur Google pour le partage de produits tels que Gmail, Photos et Docs. Les Conditions d'Utilisation de Google indiquent que Google conserve la licence d'utilisation des données stockées sous diverses formes, notamment la reproduction, la modification, la communication et la publication¹⁰⁸. Bien que toutes les données des utilisateurs stockées sur le

106 J.D. Biersdorfer, "Getting social with Waze," The New York Times, 20 septembre 2017, disponible à l'adresse sur <https://www.nytimes.com/2017/09/20/technology/personaltech/getting-social-with-waze.html>

107 "Google Cloud help, privacy," Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://support.google.com/googlecloud/answer/6056650?hl=en>

lecteur Google soient chiffrées, il ne s’agit pas d’un « chiffrement zero-knowledge »¹⁰⁹ puisque Google gère la clé de chiffrement des données.

d) Applications de tchat vidéo et de messagerie sociale

107. Google Hangouts est une plateforme de communication, semblable à Skype, elle fait partie des applications connectées au cloud de la G Suite de Google. Les utilisateurs peuvent démarrer et participer à des vidéoconférences ou à des conversations de groupe à partir de l’application Hangout disponible à l’adresse Android et iOS, via un navigateur Web, depuis l’application de bureau Hangouts ou l’extension Chrome, et depuis d’autres produits Google (par exemple Gmail et Calendar)¹¹⁰. Google stocke les détails de ces échanges (y compris l’horodatage des conférences téléphoniques et des conversations), les informations sur les participants et le contenu des messages. Ces informations sont à la disposition des utilisateurs et peuvent être téléchargées via l’outil Google Takeout¹¹¹. La figure 20 montre quelques données enregistrées lors d’une conversation Google Hangouts, , notamment les noms des participants, leurs identifiants et le contenu de la conversation.

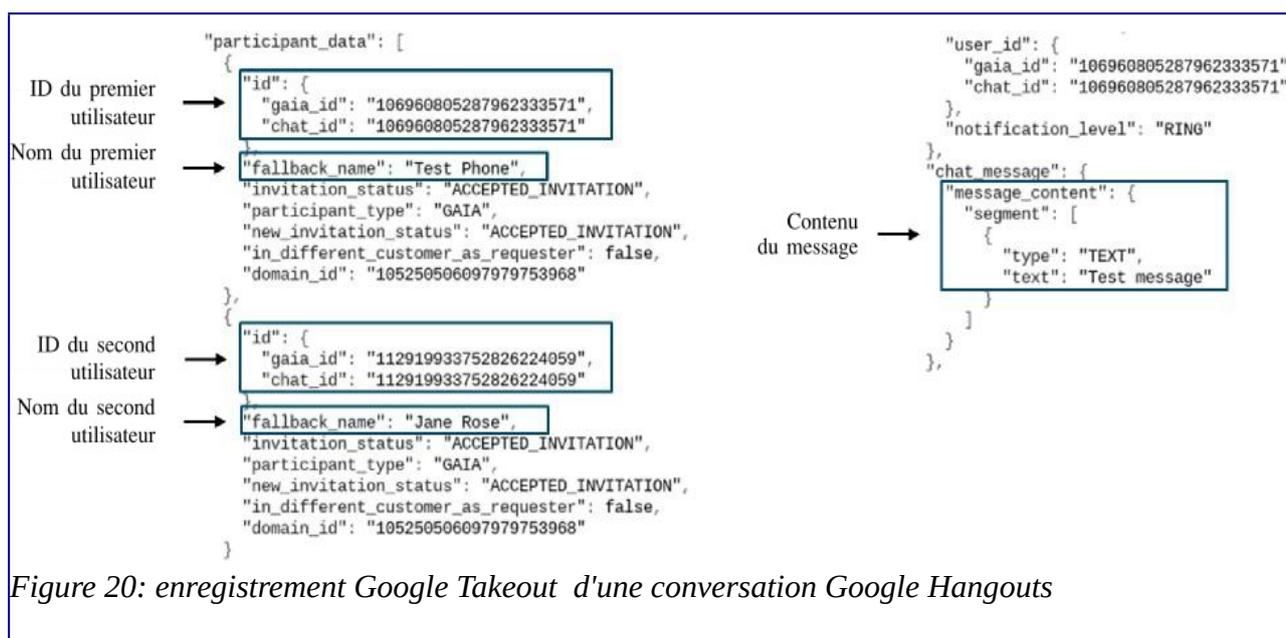


Figure 20: enregistrement Google Takeout d'une conversation Google Hangouts

108. En plus de la version Hangouts pour entreprise, Google propose également une application de messagerie instantanée appelée Allo, qui est disponible pour Android et iOS ou accessible via des navigateurs Web¹¹². Google enregistre et stocke tous les messages communiqués par Allo par défaut (sauf si l'utilisateur lance le mode incognito)¹¹³.

108 "Google Drive terms of service," Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://www.google.com/drive/terms-of-service>

109 Fergus O'Sullivan, "What exactly is zero-knowledge in the cloud and how does it work?," Cloudwards, 16 juin 2017, disponible à l'adresse <https://www.cloudwards.net/what-exactly-is-zero-knowledge-in-the-cloud-and-how-does-it-work>

110 "GSuite learning center," Google, dernière consultation le 15 août 2018, disponible à l'adresse [https://gsuite.google.com/learning-center/products/hangouts/get-started/# !!](https://gsuite.google.com/learning-center/products/hangouts/get-started/#!!)

111 "Download your data," Google, disponible à l'adresse <https://takeout.google.com/settings/takeout?pli=1>

112 Sean Keach, "Google Allo just got a major upgrade – but do we really need it?," Trusted Reviews, 16 août 2017, disponible à l'adresse <http://www.trustedreviews.com/news/google-allo-3261336>

113 Russell Brandom, "Google backs off on previously announced Allo privacy feature," The Verge, 21 septembre 2016, disponible à l'adresse <https://www.theverge.com/2016/9/21/12994362/allo-privacy-message-logs-google>

109. Google propose une application appelée Google Duo qui est une application de tchat vidéo mobile dédiée. Elle est disponible à l'adresse Android et iOS, mais pas sur les ordinateurs de bureau/ordinateurs portables. Les utilisateurs peuvent s'appeler ou tchatter en vidéo et même se connecter avec des utilisateurs du système d'exploitation Android qui n'ont pas téléchargé l'application¹¹⁴. Une fois installé, Google Duo a accès aux données du profil de l'utilisateur, aux contacts, à l'appareil photo, au microphone, aux informations de connexion Wi-Fi, à l'identifiant de l'appareil et aux informations d'appel¹¹⁵. Google stocke également les horodatages d'utilisation des applications et fournit ces informations aux utilisateurs via Google Takeout.

e) Google+

110. Google+ est un réseau social lancé en 2011 en tant que concurrent de Facebook. Bien que Google ne diffuse pas les statistiques d'utilisateurs actifs de Google+, c'est principalement un réseau pour les communautés de niche¹¹⁶. Un utilisateur peut choisir de suivre d'autres utilisateurs, d'organiser les utilisateurs qu'il suit en groupe (par ex. meilleurs amis ou collègues de travail), démarrer des « communautés », ou en rejoindre celles déjà existantes (par ex. « amateurs de photographie »). Les publications des utilisateurs suivis et des communautés apparaissent ensuite dans le flux de l'utilisateur.

111. Google Takeout montre différent type de données de Google+ que Google stocke. Google compile les vCards des profils de tous les profils qu'un utilisateur suit (dans Takeout, ces informations sont contenues dans le répertoire « Cercles Google+ »). Google collecte aussi au format HTML tout le contenu posté par un utilisateur, qui se trouve dans le dossier « Flux Google+ » de Takeout, comme montré sur la figure 21. Les photos postées sont aussi sauveées dans Google Photos.



Figure 21: Exemple du contenu d'une publication Google+ sauvegardée par Google

f) Translate

112. Google traduction est un service gratuit de traduction automatique qui supporte plus de 100 langues, qui est accessible sur le web et à travers des applications Android et iOS. Il est aussi intégré à l'assistant Google et à Google Chrome, et est disponible pour les développeurs d'applications tiers via une API payante¹¹⁷. Tout compris, il sert plus de 500 millions d'utilisateurs mensuels¹¹⁸. Si les utilisateurs ont l'application Google traduction sur leur téléphone, ils peuvent l'utiliser pour la

114 Swapna Krishna, "Google Duo allows you to call people who don't have the app," Engadget, 12 janvier 2018, disponible à l'adresse <https://www.engadget.com/2018/01/12/google-duo-call-android-users-without-app>

115 Google informe l'utilisateur de cet accès lorsque l'utilisateur télécharge l'application

116 Karissa Bell, "Google+ isn't dead and these are the people still using it the most," Mashable, 18 janvier 2017, disponible à l'adresse https://mashable.com/2017/01/18/who-is-using-google-plus-anyway/#_lPOPXr.eiqZ

117 "Google Cloud, pricing," dernière consultation le 15 août 2018, disponible à l'adresse <https://cloud.google.com/translate/pricing>

118 Gideon Lewis-Kraus, "The great A.I. awakening," The New York Times, 14 décembre 2016, disponible à l'adresse <https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html>

traduction dans d'autres applications, comme WhatsApp¹¹⁹. Bien que Google affirme ne pas suivre les requêtes web de traduction¹²⁰ (et que ces recherches n'apparaissent pas dans l'historique de recherche ou ailleurs dans Google Takeout), la politique de confidentialité de Google l'autorise à les stocker pour une courte période pour déboguer ou pour d'autres tests¹²¹.

g) Calendar

113. Google Calendar est un outil de planification qui permet aux utilisateurs de garder une trace de leurs activités quotidiennes et hebdomadaires. Il est largement utilisé sur les ordinateurs fixes et sur les appareils mobiles, avec plus de 500 millions de téléchargements dans le Google Play Store¹²². Des informations personnelles, comme le nom d'une personne et ses informations de contact, sont souvent associées à l'utilisation de Calendar.

114. En utilisant les calendriers, les utilisateurs fournissent à Google des détails (tels que l'heure, l'endroit et les informations de contact) de tous les participants d'un événement. Google indique que l'application a accès à la « lecture des événements du calendrier mais aussi à des informations confidentielles »¹²³. Grâce à cela, Google lit et stocke les adresses électroniques de tous les participants à un événement du calendrier, que ce soit des courriels Google ou non¹²⁴. Dès qu'une personne participant à l'événement utilise l'application calendrier de Google, Google enregistre les informations personnelles de tous les participants.

h) Keep

115. Google Keep est un outil de prise de note qui permet à l'utilisateur de saisir des pense-bêtes qui se synchronisent entre les différents appareils qui sont associés à son compte. Keep a été téléchargé plus de 100 millions de fois depuis le Google Play Store¹²⁵. Google collecte et stocke l'ensemble du contenu des notes ainsi que l'heure de leur création¹²⁶. La figure 22 montre des notes de Google Keep enregistrées par Google. Google analyse et classe les notes qui sont créées selon leur contenu. Par exemple, nourriture, lieux et voyage sont des catégories de classification.

119 "Google Translate," Google, disponible à l'adresse <https://translate.google.com/intl/en/about>

120 "Google Translate help," Google, dernière consultation le 15 août 2018, disponible à l'adresse <https://support.google.com/translate/answer/6142479?co=GENIE.Platform%3DDesktop&hl=en&oco=0>

121 "Access to the Google Cloud API," Google, dernière consultation le 15 août 2018, disponible à l'adresse <https://cloud.google.com/translate/faq>

122 "Google Calendar," Google Play Store, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://play.google.com/store/apps/details?id=com.google.android.calendar>

123 "Google Calendar," Google Play Store, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://play.google.com/store/apps/details?id=com.google.android.calendar>

124 "Download your data," Google, disponible à l'adresse <https://takeout.google.com/settings/takeout?pli=1>

125 "Google Keep," Google Play Store, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://play.google.com/store/apps/details?id=com.google.android.keep&hl>

126 "Download your data," Google, disponible à l'adresse <https://takeout.google.com/settings/takeout?pli=1>

```

</style></head>
<body><div class="note"><div class="heading"><div class="meta-icons">
</div>
Apr 26, 2018, 11:06:04 AM</div>
<div class="content">Grocery list:<br>- pasta<br>- tomatoes<br>- lettuce <br>- salad
dressing <br>- salmon <br>- lemon<br>- capers <br>- Bread<br>- butter<br>- milk<br>-
eggs<br>- bacon<br>- OJ<br>- </div>
</div></body></html>

```

Figure 22: Exemple d'une note Google Keep, récupérée dans Google take out

i) Chromecast

116. Tout comme Apple TV, Google Chromecast est un appareil qui sert d'interface pour regarder des vidéos au travers de diverses applications (i.e. Netflix, YouTube, Hulu, Play Store), mais aussi de projeter des vidéos depuis les smartphones ou les ordinateurs sur des écrans ou des télévisions plus grands. Tous les appareils de diffusion ont des clefs d'identification qui sont associées à un compte Google au moment de leur enregistrement.

117. Google utilise Chromecast pour agréger les activités du système, les rapports d'erreurs, les données d'utilisation et les détails sur les fonctionnalités de diffusion utilisées sur les appareils. Cela inclut les applications et les noms de domaine qui sont projetés¹²⁷. ChromeCast utilise Google Cast qui est une plateforme logicielle permettant la diffusion du son et de l'image entre des appareils d'un même réseau¹²⁸. En plus d'une multitude de produits Google (par exemple Google Home) qui utilise cette fonctionnalité, la plateforme est aussi utilisée par des produits tiers qui intègrent la technologie ChromeCast. (comme des Télévisions des consoles de jeux vidéos¹²⁹) qui offrent une fonctionnalité similaire. Google collecte également des données d'utilisation et d'erreurs de ces appareils de diffusion tiers¹³⁰.

j) Google DNS

118. Google a lancé en décembre 2009 un service gratuit de résolution de nom de domaine. (DNS)¹³¹ appelé Google Public DNS. Le DNS de Google a pour objectif de fournir une navigation améliorée en terme de rapidité, sécurité et précision¹³². En décembre 2014, Google Public DNS aurait fournit 400 milliards d'adresses IP par jour¹³³.

127 "Chromecast help", Google, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://support.google.com/chromecast/answer/6076570?hl=fr>

128 "Google Cast," Google, dernier accès le 15 août 2018, disponible à <https://developers.google.com/cast>

129 "Qu'est ce que Google Cast and Chromecast," Shield, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://shield.nvidia.com/blog/what-is-googlecast-chromecast>

130 "Aide ChromeCast" Google, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://support.google.com/chromecast/answer/6076570?hl=fr>

131 Les services DNS traduisent les noms de domaine en adresses IP et sont à ce titre nécessaires pour naviguer sur le Web.

132 "Google Public DNS", Google, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://developers.google.com/speed/public-dns/faq>

133 "Google Public DNS et les réponses dépendantes de la localisation de l'utilisateur" Google Webmaster Central Blog, 15 décembre 2014, disponible à l'adresse <https://webmasters.googleblog.com/2014/12/google-public-dns-and-location.html>

119. Pour détecter les abus (comme les attaques DDoS) et pour apporter des solutions, Google Public DNS conserve un log temporaire des adresses IP complètes qu'il efface dans un délai de 24 à 48 heures. Pour des besoins de *debug* et de préventions des abus nécessitant plus de temps, Google conserve des données anonymisées à une granularité de la ville pendant 2 semaines et garde un sous-ensemble aléatoire de ces données de façon permanente¹³⁴.

k) Routeur Wi-Fi Google

120. En décembre 2016, Google a commencé à déployer des routeur Wi-Fi maillés. Les routeurs maillés permettent à un utilisateur d'étendre la portée d'un réseau Wi-Fi en ajoutant des routeurs connectés entre eux. En décembre 2017, Google Wi-Fi est devenu le meilleur système Wi-Fi maillé des États-Unis selon les données de NPD Group¹³⁵. Les informations collectées¹³⁶ se répartissent dans ces trois catégories :

- Services hébergés, ce qui inclut des informations diffusées par les appareils connectés (comme le nom de l'appareil « Jane's iPhone »), informations déduites de ces appareils (tel que la marque, « Samsung »), le statut de connexion, la vitesse de transfert des données, les rapports d'erreurs et la quantité de données consommée
- Statistiques Wi-Fi, ce qui inclut l'utilisation de données anonymisée, rapports d'erreurs et des informations sur les performances des appareils
- Statistiques des applications Wi-Fi, tels que l'utilisation et les rapports d'erreur

121. D'après Google, le routeur Wi-Fi n'enregistre pas les pages visitées ni ne collecte le contenu du trafic réseau¹³⁷

I) Nest

122. En Janvier 2014, Google a racheté Nest, une société de domotique¹³⁸. La gamme de produits Nest comprend de nombreux objets électroménagers connectés, comme des thermostats, des caméras, des sonnettes, des systèmes d'alarmes et des détecteurs de fumée¹³⁹. Il existe également un liste de plus de 100 produits partenaires qui s'intègrent avec Nest, qui vont des réfrigérateurs aux lits¹⁴⁰.

123. Les objets Nest collectent des informations variées, qui comprennent non seulement les paramétrages directs des utilisateurs sur les appareils, mais aussi les données de l'environnement interne de la maison. Par exemple, le « Nest Learning Thermostat » collecte des données (comme la température, l'humidité, la lumière ambiante et le mouvement) ce qui lui permet de savoir quand les

134 "Google Public DNS," Google, disponible à l'adresse <https://developers.google.com/speed/public-dns/privacy>

135 Jillian D'Onfro, "The surprising use case that has made Google Wifi one of the company's sleeper hits" CNBC, 18 décembre 2017, disponible à l'adresse <https://www.cnbc.com/2017/12/18/google-wifi-mesh-technology-sales-stats.html>

136 "Google Wifi help", Google, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://support.google.com/wifi/answer/6246642?hl=fr>

137 "Google Wifi help", Google, visité la dernière fois le 15 août 2018, disponible à l'adresse <https://support.google.com/wifi/answer/6246642?hl=fr>

138 "Google to acquire Nest", Alphabet, 13 janvier 2014, disponible à l'adresse <https://abc.xyz/investor/news/releases/2014/0113.html>

139 Page d'accueil de Nest : <https://nest.com/>

140 "Works with Nest," Nest, page en français ici : <https://nest.com/fr/works-with-nest/>

gens sont à la maison et même dans quelle pièce¹⁴¹. Quand un utilisateur connecte un outil tiers compatible avec Nest, Nest partage des informations avec ces tiers mais informe l'utilisateur des informations qui vont être partagées¹⁴². Nest ne partage pas de données avec d'autres acteurs tels que qu'un fournisseur d'énergie ou une assurance partenaire sans obtenir le consentement de l'utilisateur¹⁴³.

124. Sur son site Internet, Nest indique que les comptes Nest et Google ne sont pas liés (à moins qu'un utilisateur choisisse d'intégrer les produits et services Google) et que Google ne vend pas les données de Nest. Mais récemment, le partage de données a suscité des inquiétudes suite à une publicité¹⁴⁴ de Google sur le fusionnement des équipes de matériel de Nest et Google¹⁴⁵. De plus, les inquiétudes concernant les futurs liens entre Google, Nest et d'autres compagnies d'électricité et d'assurance sont apparues dans les médias¹⁴⁶.

m) Google Fiber

125. Google Fiber c'est de l'internet haut débit, une télévision numérique (TV sur IP) et un service de téléphonie VOIP (voix sur IP) qui connectent les utilisateurs à un réseau de fibres optiques à très haute vitesse qui couvrent toute l'habitation¹⁴⁷ qui s'appellent respectivement, Fiber Internet, Fiber TV et Fiber Phone. Les efforts de Google pour déployer un réseau étendu de fibre optique ont été ralenti par des coûts matériels et des négociations avec les municipalités locales. C'est pourquoi l'extension à de nouvelles villes passe maintenant par Webpass (un fournisseur d'accès racheté par Google), qui fournit des débits équivalents avec des technologies sans fil et ethernet existantes¹⁴⁸

126. Fiber Internet collecte des données techniques sur le compte Google d'un utilisateur, mais ne partage pas les détails du compte avec d'autres propriétés de Google sans consentement supplémentaire de l'utilisateur¹⁴⁹. Fiber Internet nécessite le consentement de l'utilisateur avant d'associer le compte Google d'un utilisateur avec d'autres informations, comme des sites visités ou le contenu de communications¹⁵⁰.

141 "Privacy statement for Nest products and services," Nest, accédé pour la dernière fois le 15 août 2018, disponible à l'adresse <https://nest.com/legal/privacy-statement-for-nest-products-and-services/>

142 "Privacy statement for Nest products and services," Nest, accédé pour la dernière fois le 15 août 2018, disponible à l'adresse <https://nest.com/legal/privacy-statement-for-nest-products-and-services/>

143 "Privacy statement for Nest products and services," Nest, accédé pour la dernière fois le 15 août 2018,, disponible à l'adresse <https://nest.com/legal/privacy-statement-for-nest-products-and-services/>

144 Rick Osterloh, "Nest to join forces with Google's hardware team," Nest, 7 février 2018, disponible à l'adresse <https://blog.google/topics/hardware/nest-join-forces-googles-hardware-team/>

145 Leo Kelion, "Google-Nest merger raises privacy issues," BBC, 8 février 2018, disponible à l'adresse <http://www.bbc.com/news/technology-42989073>

146 Casey Johnston, "What Google can really do with Nest, or really, Nest's data," Ars Technica, 15 janvier 2014, disponible à l'adresse <https://arstechnica.com/information-technology/2014/01/what-google-can-really-do-with-nest-or-really-nests-data/>

147 Ryan Waniata, "Comcast killer : Google Fiber touches down in Austin with its new TV and internet devices," Digital Trends, 3 décembre 2014, disponible à l'adresse <https://www.digitaltrends.com/home-theater/google-fiber-tv-hands-on>

148 Nick Statt, "Google Fiber-owned Webpass is bringing its wireless gigabit internet to Denver," The Verge, 22 février 2017, disponible à l'adresse <https://www.theverge.com/2017/2/22/14703142/google-fiber-webpass-denver-expansion-gigabit-Internet> and <https://gizmodo.com/what-happened-to-google-fiber-1792440779>

149 "Google Fiber privacy notice," Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://fiber.google.com/legal/privacy>

150 "Google Fiber privacy notice," Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://fiber.google.com/legal/privacy>

127. Fiber TV collecte des informations sur les programmes et les applications utilisés et les associe au compte Google de l'utilisateur¹⁵¹. De même, Fiber Phone collecte les informations d'utilisation (par ex. les logs de l'historique d'appels, messages vocaux sur répondeur, messages SMS, et conversations enregistrées) et les associe au compte Google de l'utilisateur¹⁵². Bien que ces informations ne soient pas partagées à des tiers sans le consentement de l'utilisateur, elles le sont pour du traitement externe ou des raisons juridiques. Alors que Google ne fait pas référence de manière explicite à l'utilisation de Fiber TV/phone pour le ciblage publicitaire, sa politique de confidentialité générale autorise néanmoins l'utilisation des informations personnelles collectées à des fins de ciblage. De plus, Google affirme partager les informations identifiables et non-personnelles des utilisateurs récoltées par Google Fiber publiquement aux fournisseurs de contenu, éditeurs, publicitaires, et/ou sites connectés¹⁵³.

D. Méthode du suivi de trafic selon la localisation

128. Pour capter les requêtes envoyées vers les domaines des serveurs de Google depuis un téléphone portable, notre étude utilise la technique du "man-in-the-middle" (MITM) à l'aide d'un outil "Proxy MITM". Alors que les études précédentes, qui ont analysé les données similaires, se servaient d'un hotspot Wi-Fi comme intermédiaire entre le téléphone portable et les serveurs de Google, la présente étude utilise un VPN (Virtual Private Network) sur le téléphone portable pour analyser les données envoyées par les réseaux Wi-Fi ainsi que les réseaux mobiles.

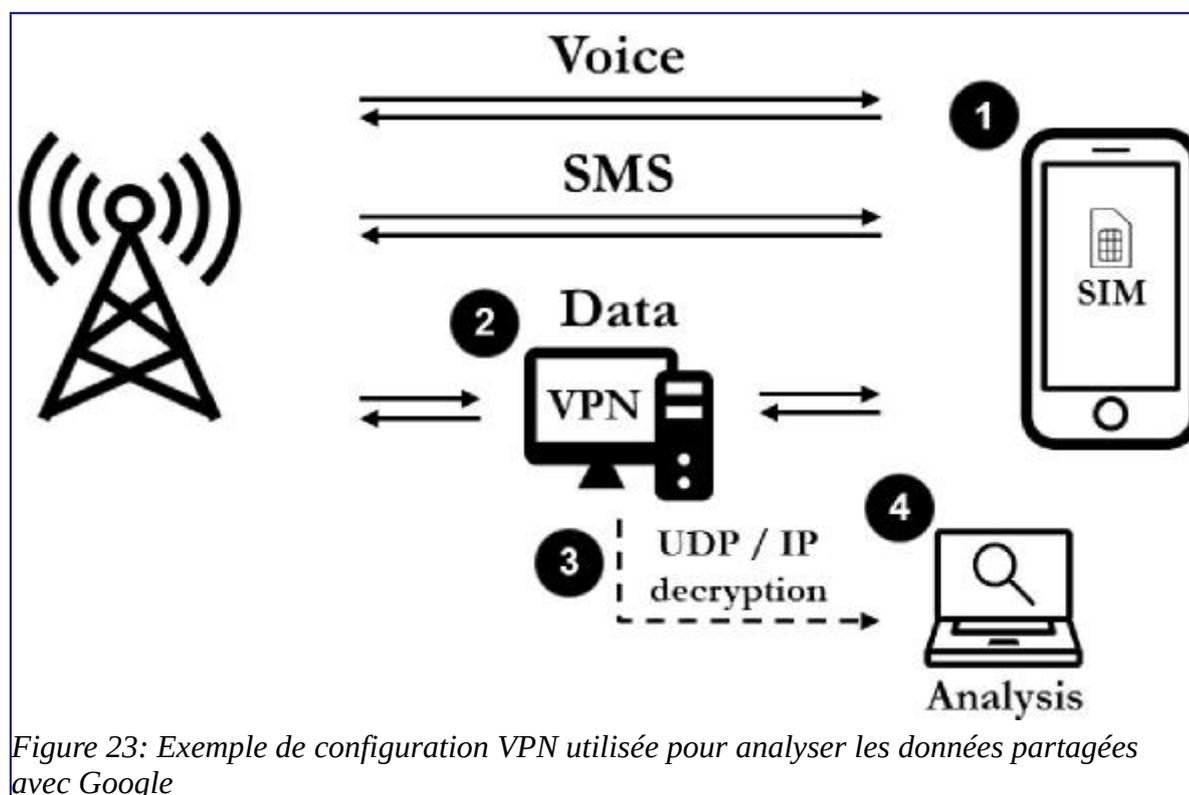


Figure 23: Exemple de configuration VPN utilisée pour analyser les données partagées avec Google

151 "Google Fiber privacy notice," Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://fiber.google.com/legal/privacy>

152 "Google Fiber privacy notice," Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://fiber.google.com/legal/privacy>

153 "Google Fiber privacy notice," Google, visité pour la dernière fois le 15 août 2018, disponible à l'adresse <https://fiber.google.com/legal/privacy>

Voici ci-dessous les étapes spécifiques qui ont été suivies pour configurer et mener les tests :

1. Les appareils mobiles utilisés dans les tests de collecte de données¹⁵⁴ ont été réinitialisés aux paramètres d'usine pour s'assurer qu'aucune application installée au préalable ou qu'aucun réglage de paramètre n'affecte le trafic vers et depuis le téléphone. Lors de la réactivation, les appareils sont configurés selon les paramètres par défaut suggérés. Ensuite, on insère de nouvelles cartes SIM dans les appareils pour obtenir de nouveaux numéros de téléphone portable.
2. Une connexion VPN est configurée avec un ordinateur proxy à distance à l'aide des paramètres/fonctionnalités du VPN fournis par le système d'exploitation du téléphone. On a choisi un IPsec/L2TP avec un mode d'authentification PSK. La configuration VPN permet au proxy à distance d'intercepter et d'enregistrer les données transmises depuis les téléphones portables. En raison de la nature du type de signaux émis depuis un téléphone, le réglage du VPN ne peut pas intercepter les voix et les données SMS depuis les appareils portables. Cependant, il capte bien tout le trafic TCP depuis et vers l'appareil, notamment le trafic HTTP et HTTPS.
3. Des certificats logiciels HTTPS ont été installés sur les appareils portables pour permettre de décrypter le trafic de données capté. La configuration VPN permet le routage du trafic HTTP et HTTPS via le logiciel mitmproxy¹⁵⁵ à l'aide d'iptables¹⁵⁶. Ce programme procède au décryptage SSL en utilisant son propre certificat pour décrypter le trafic et le déposer dans des fichiers HAR (Archive HTTP) pour analyse.
4. L'analyse des données de trafic HTTP and HTTPS décryptées consiste principalement à catégoriser les requêtes de serveur en segments clés grâce aux informations d'en-tête des requêtes. Les Tableaux 5 et 6 montrent les en-têtes de trafic qui ont été identifiés comme données en cours de transmission vers/depuis Google et Apple.

129. Dans des cas précis, les requêtes à Google ont été davantage décodées pour analyser les informations transmises à un niveau plus granulaire. Une requête spécifique à Google qui a été davantage décodée était la « Google location API », désignée par le paramètre `/loc/m/api`. Les spécifications de localisation ont été analysées en retirant l'en-tête du message et en décodant le message Protobuf compressé¹⁵⁷. L'API de localisation décodée contenait les analyses Wi-Fi et réseau qui étaient utilisées pour déterminer la localisation de l'appareil.

154 Les appareils incluent un LG X Power sous Android 6.0 et un iPhone 5 sous iOS 10.3.3

155 <https://mitmproxy.org/>

156 "iptables(8) – Linux man page," Die.net, disponible à l'adresse <https://linux.die.net/man/8/iptables>

157 Des informations supplémentaires sur les méthodes de décodage peuvent être trouvées ici : "Reverse engineering : Google Location protobuf specifications," Esther Codes, lu en mars 2018, disponible à l'adresse <https://web.archive.org/web/20180213201547/https://esther.codes/reverse-engineering-google-location-gms-specification/>

Segment	En-tête de routage	Description
Domaines publicitaires	doubleclick.net	envoie et reçoit des données de doubleclick.net
	google-analytics.com	envoie et reçoit des données de GA
	googletagmanager.com	insère des balises HTML
	googletagservices.com	insère des balises HTML
	googlesyndication.com	télécharge et affiche des publicités
	adservice.google.com	appelle le réseau adWords
	google.com/ads/	comptabilise les annonces et les utilisateurs
	gstatic.com	charge des publicités sur les pages
	google.com/adsense	appelle adSense
	google.com/pagead	redirige vers les pages pointées par les publicités
Localisation	google.com/loc/m/api	retourne les informations sur les réseaux et Wi-Fi aux alentours
	googleapis.com/userlocation/v1/reports	retourne les informations de déplacement de l'utilisateur (tel que marche, course, vélo, conduite...)
	googleapis.com/placesandroid	associe des positions avec des commerces (google places)
	maps.googleapis.com/maps	récupère les informations détaillées sur un lieu identifié ou recherché
	clients4.google.com/glm/mmap	envoie les informations de géolocalisation d'un utilisateur pour récupérer les données de cartographie
API Google Play	play.googleapis.com/log/batch	informations sur l'activité du périphérique
	play.googleapis.com/play/log/timestamp	met à jour un cookie et retourne la date et l'heure
	play.googleapis.com/play/log?format	envoie les informations sur le play Store
Device auth. et upload	googleapis.com/batch	informations sur le périphérique et les mises à jour
	clients4.google.com/chrome-sync	synchronisation du navigateur Chrome
	googleapis.com/experimentsandconfigs	teste le téléchargement de configuration
	android.clients.google.com/backup/backup	sauvegarde du périphérique

	android.clients.google.com/auth	autorisation du périphérique
	android.clients.google.com/checkin	enregistrement et identification du périphérique
	android.clients.google.com/	authentification aux services Google
Autres	cdn.ampproject.org	récupération de données sur le réseau de distribution de contenu AMP de Google
	google.com/xjs	échange de données avec Google Search
	google.com/gen	échange de données avec Google Search et transmission d'un cookie
	doubleclick.net	envoie et reçoit des données de doubleclick.net
	google-analytics.com	envoie et reçoit des données de GA
	googletagmanager.com	insère des balises HTML
	googletagservices.com	insère des balises HTML
	googlesyndication.com	télécharge et affiche des publicités
	adservice.google.com	appelle le réseau adWords

Tableau 5 : Noms de domaines de Google remarquables communiquant avec des téléphones mobiles

Segment	En-tête de routage	Description
Localisation	geosrc=wifi,73.xxx&kb_ime=en_US&key=beagle1626&latlng=40.xxx,-74.xxx&locale=en_US	Renvoie les coordonnées de localisation à Apple lorsque le navigateur est utilisé
	cl2.apple.com	Appel le serveur de localisation principal
	gs-loc.apple.com	Services de localisation Apple
App store	us-east-1.blobstore.apple.com/apple	Communique avec l'Apple store ; inclut Apple ID
Authentification de l'appareil et envoi	mesu.apple.com/assets/com_apple_MobileAsset	Détaille la configuration et les paramètres du mobile
	bookmarks.icloud.com	Synchronise le comportement du mobile avec le cloud
	ckdatabase.icloud.com/api	Communique le jeton d'autorisation de l'appareil et synchronise avec iCloud
	keyvalueservice.icloud.com/sync	Synchronise le comportement du mobile
Autres	api-glb-nyc.smoot.apple.com/	Diverses APIs Apple
	gsp64-ssl.ls.apple.com	Fournit les informations de l'appareil quand le téléphone accède aux sites internet via Safari
	gspe35-ssl.ls.apple.com/geo_manifest/dynamic/config?application=geod	Charge les tuiles de cartes, mais ne transmet pas d'informations de localisation
	configuration.apple.com/configurations/pep/config/geo/networkDefaults	Communique les paramètres des outils de collecte de localisation

Tableau 6 : noms de domaines Apple remarquables communiquant avec les appareils de téléphonie mobile

E. Séquence d'authentification de l'inscription à Google

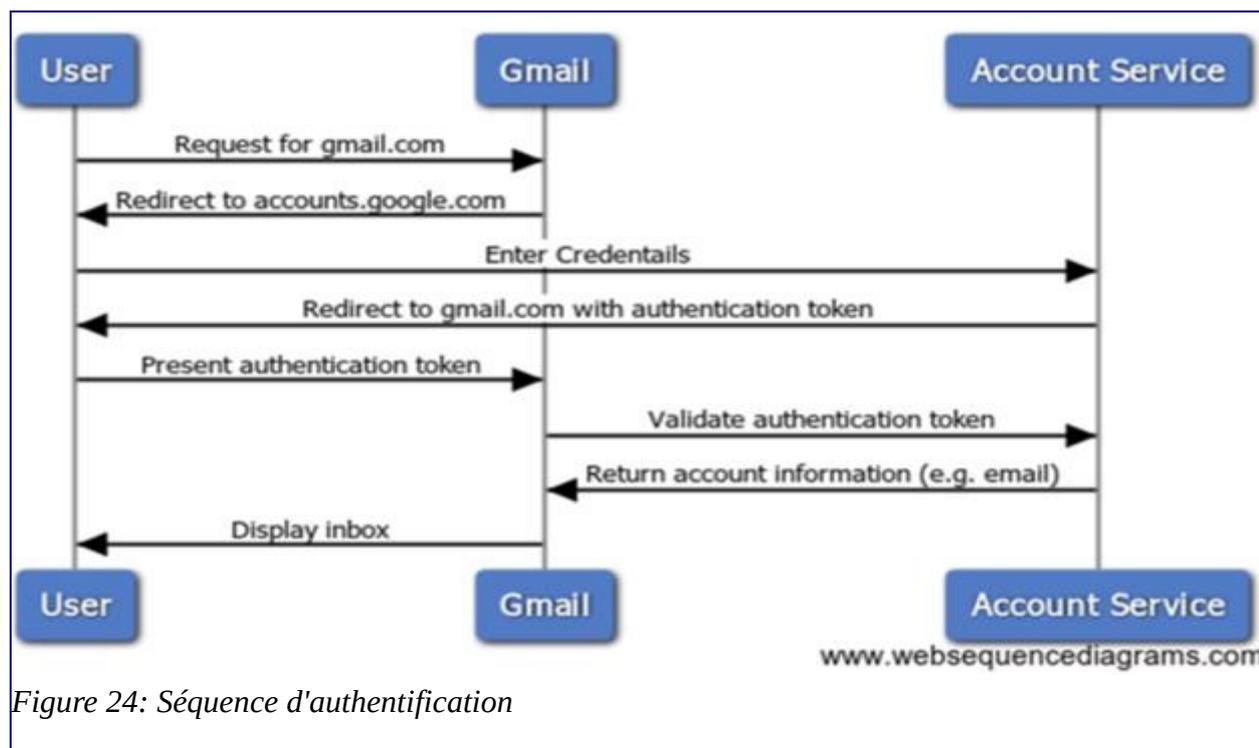


Figure 24: Séquence d'authentification

F. Profil d'utilisation pour les expériences de collecte de données du mobile

130. Un profil d'utilisation a été conçu pour simuler l'interaction d'un utilisateur normal avec son téléphone mobile au cours de la journée. Diverses statistiques décrivant le comportement en ligne des populations ont été intégrées pour créer le profil, comme décrit en dessous.

131. Le profil conçu a visité 45 pages internet au cours de la journée, inspiré par les statistiques de visite des sites internet d'une étude de Nielsen de 2010 qui indiquent que la personne moyenne visite 88 pages internet par jour¹⁵⁸ et une étude de Stone Temple de 2017 qui indique qu'environ 50 à 55 % des visites de page internet viennent d'appareils mobiles¹⁵⁹. Les 45 visites de page internet ont été uniformément réparties entre 5 grands sites d'actualité et de sport hors Google¹⁶⁰. Lorsque le téléphone n'était pas en cours d'utilisation pour visiter des pages internet, le navigateur continuait de tourner en arrière-plan. Le profil d'utilisation qui en résulte représente un utilisateur [prudent / conservateur] puisque le nombre de visites quotidiennes de page internet est susceptible d'avoir augmenté depuis l'étude de Nielsen de 2010.

132. Le profil d'utilisation comprenait aussi une variété d'applications mobiles non-Google. Les applications phares hors Google ont été sélectionnées dans les catégories des médias sociaux, du shopping, du voyage et de la santé. Ces applications incluaient Facebook, Instagram, Snapchat, Pinterest, Amazon Shopping, Walmart, Starbucks, Yelp, et Six Pack in 30 Days. Ces applications ont

158 "Nielsen provides topline U.S. web data for March 2010," Nielsen, avril 2010, disponible à l'adresse <http://www.nielsen.com/us/en/insights/news/2010/nielsen-provides-topline-u-s-web-data-for-march-2010.html>

159 Eric Enge, "Mobile vs desktop usage : mobile grows but desktop still a big player in 2017," Stone Temple, avril 2017, disponible à l'adresse <https://www.stonetemple.com/mobile-vs-desktop-usage-mobile-grows-but-desktop-still-a-big-player-in-2017/>

160 Les sites sélectionnés étaient New York Time, CNN, The Guardian, ESPN, et Crickbuzz. Les sites internet ont été déterminés en utilisant la liste d'Alexa, disponible à l'adresse <https://www.alexa.com/topsites/category>

été ouvertes périodiquement au cours de la journée pour simuler un utilisateur typique qui passe environ 2 heures et demie par jour sur des applications mobiles, comme publié par eMarketer¹⁶¹.

G. Articles publiés sur les pratiques de collecte de données mises en œuvre par Google

Titre	Point principal	Auteur, date
Exclusivité AP : Google trace vos mouvements, que vous le vouliez ou non ¹⁶²	Google trace la localisation des utilisateurs même lorsque les services de localisation sont désactivés	Ryan Nakashima, août 2018
Une autorité australienne enquête sur la récolte de données par les appareils Android ¹⁶³	Google « récolte » environ 1 Go de données par mois depuis des appareils Android	Oracle, mai 2018
Comment empêcher Google de posséder votre vie en ligne ¹⁶⁴	Il est très difficile pour le consommateur moyen d'éviter les produits de Google	WSJ, mai 2018
Google trace les téléphones même lorsqu'ils sont déconnectés ? ¹⁶⁵	Google trace les téléphones même lorsque les téléphones sont « déconnectés » (pas de carte SIM, mode avion, Wi-Fi éteint)	Fox News, février 2018
Google collecte la localisation des utilisateurs d'Android même lorsque les services de localisation sont désactivés ¹⁶⁶	Google collecte la localisation d'Android lorsque les services de localisation sont coupés	Quartz, novembre 2017
Suivi en ligne : Une mesure et analyse d'un million de sites ¹⁶⁷	Google peut suivre des utilisateurs d'environ 80 % des sites internet grâce à ses cookies.	Princeton University, 2016

161 "eMarketer reveals new estimates for mobile app usage," eMarketer, avril 2017, disponible à l'adresse <https://www.emarketer.com/Article/eMarketer-Unveils-New-Estimates-Mobile-App-Usage/1015611>

162 Ryan Nakashima, "AP Exclusive : Google tracks your movements, like it or not," AP, 13 août 2018, disponible à l'adresse <https://apnews.com/828aefab64d4411bac257a07c1af0ecb>

163 Anne Davis, "Australian regulator investigates Google data harvesting from Android phones," The Guardian, 13 mai 2018, disponible à l'adresse <https://www.theguardian.com/technology/2018/may/14/australian-regulator-investigates-google-data-harvesting-from-android-phones>

164 David Pierce, "How to Keep Google From Owning Your Online Life," The Wall Street Journal, 8 mai 2018, disponible à l'adresse <https://www.wsj.com/articles/how-to-keep-google-from-owning-your-online-life-1525795372>

165 Brett Larson, « Google tracking phones even when they are disconnected ?, » Fox News, 11 février 2018, disponible à l'adresse <http://video.foxnews.com/v/5731183327001/?#sp=show-clips>

166 Keith Collins, "Google collects Android users' locations even when location services are disabled," Quartz, 17 novembre 2017, disponible à l'adresse <https://qz.com/1131515/google-collects-android-users-locations-even-when-location-services-are-disabled/>

167 Englehardt, Steven, et Arvind Narayana, "Online Tracking : A 1-million-site Measurement and Analysis," ACM CCS, 2016, disponible à l'adresse http://randomwalker.info/publications/OpenWPM_1_million_site_tracking_measurement.pdf

Pourquoi les Smartphones Android dévorent le plus les données ? ¹⁶⁸	Les appareils Android consomment plus de données (2,2 Go/mois) que les autres smartphones	Ericsson, décembre 2013
Fuite de données depuis les smartphones Android ¹⁶⁹	Android transmet des identifiants anonymes ainsi que les identifiants de périphériques comme l'adresse Mac et l'IMEI	Lasse Øverlier, juin 2012

Tableau 7 : Résumé des autres études sur la collecte des données par Google

H. Clarifications

133. Notre compréhension des données envoyées à Google via son système Android se limite uniquement à la version 6.0. Cette étude ne tient pas compte des mises à jour/patches qui ont pu être installés sur les versions successives et qui ont pu affecter les communications d'Android avec les serveurs de Google. Alors que de nouvelles versions d'Android sont actuellement disponibles sur le marché, Android 6.0 reste encore la version la plus utilisée¹⁷⁰. De plus, même si nous avons pris le maximum de précaution pour classer les en-têtes de routage selon leur objectif (par ex. localisation, publicité, transferts vers l'appareil, app store), il est possible que certains en-têtes puissent avoir plusieurs objectifs (par ex. publicité et localisation). Ces aspects ne sont pas pris en compte dans notre étude. Donc, la description présentée pour ces en-têtes n'est sans doute pas exhaustive par rapport à l'objectif qu'ils visent.

I. À propos de l'auteur

134. Douglas Schmidt est professeur et expert en système logiciel, il a passé plus de 30 ans à conduire, superviser et faire des recherches sur le développement de logiciel pour les systèmes middleware distribués et leurs applications sur les sujets du réseau et de la sécurité, l'apprentissage machine, les smart-grids, les schémas de conception et bien d'autres. Il a écrit une dizaine de livres et plus de 600 articles qui ont été cités plus de 38 000 fois collectivement. Douglas Schmidt possède plus de 30 ans d'expérience dans l'enseignement de ces concepts en classe et en ligne à plus de 200 000 élèves au total.

135. Douglas Schmidt a collaboré à plus d'une vingtaine de missions d'expertise préalables, allant de la production de rapports d'experts à la présentation de témoignages oraux, tant lors de la déposition que lors du procès. Ses travaux de consultation vont des litiges sur les brevets et droits d'auteurs aux conseils à des entités publiques et privées sur divers problèmes concernant l'infrastructure et la conception d'un logiciel. Il a obtenu son Master et son doctorat en informatique à l'Université de Californie, Irvine en 1990 et 1994, respectivement.

168 Brian Chen, "Why Do Android Smartphones Guzzle the Most Data ?," The New York Times, 31 décembre 2013, disponible à l'adresse <https://bits.blogs.nytimes.com/2013/12/31/why-do-android-smartphones-guzzle-the-most-data/>

169 Lasse Øverlier, "Data leakage from Android smartphones," Norwegian Defense Research Establishment, 6 juin 2012, disponible à l'adresse <https://www.ffi.no/no/Rapporter/12-00275.pdf>

170 "Mobile & Tablet Android Version Market Share Worldwide," statcounter, disponible à l'adresse <http://gs.statcounter.com/android-version-market-share/mobile-tablet/worldwide>

136. Douglas Schmidt est actuellement Professeur Cornelius Vanderbilt d'Informatique à l'Université de Vanderbilt. Avant Vanderbilt, il a occupé plusieurs postes de président et directeur de service dans les secteurs académiques et industriels notamment au Software Engineering Institute de l'Université de Carnegie Mellon, à l'Information Technology Office au Defense Advanced Research Project Agency (DARPA), au gouvernement fédéral, ainsi qu'au sein de plusieurs start-ups technologiques.

J. Crédits de traduction

L'équipe Framalang par ordre alphabétique :

Alain, Barbara, Bullcheat, Côme, David_m, fab, Fabrice, FranBAG, Goofy, jums, Khrys, Mika, Obny, Penguin, Piup, Serici.

Remerciements particuliers à Cyrille.