Les données que récolte Google – Ch.3

Voici déjà la traduction du troisième chapitre de Google Data Collection, l'étude élaborée par l'équipe du professeur Douglas C. Schmidt, spécialiste des systèmes logiciels, chercheur et enseignant à l'Université Vanderbilt. Si vous les avez manqués, retrouvez les chapitres précédents déjà publiés.

Il s'agit aujourd'hui de mesurer ce que les plateformes les plus populaires recueillent de nos smartphones

Traduction Framalang : Côme, goofy, Khrys, Mika, Piup. Remerciements particuliers à badumtss qui a contribué à la traduction de l'infographie.

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 %)[1] et, en mai 2017, il y avait plus de 2 milliards d'appareils Android actifs par mois dans le monde.[2]

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 10K de 2017[3]. Les deux plateformes facilitent l'usage de contenus de Google et de tiers (p.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[4], 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 gamme 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.[5] 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.[6]

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 extensions.[7]

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 installées[8]. 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 alentours[9]. Les téléphones Android peuvent aussi utiliser les informations des balises Bluetooth enregistrées dans l'API Proximity Beacon de Google[10]. 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 immeuble.[11]

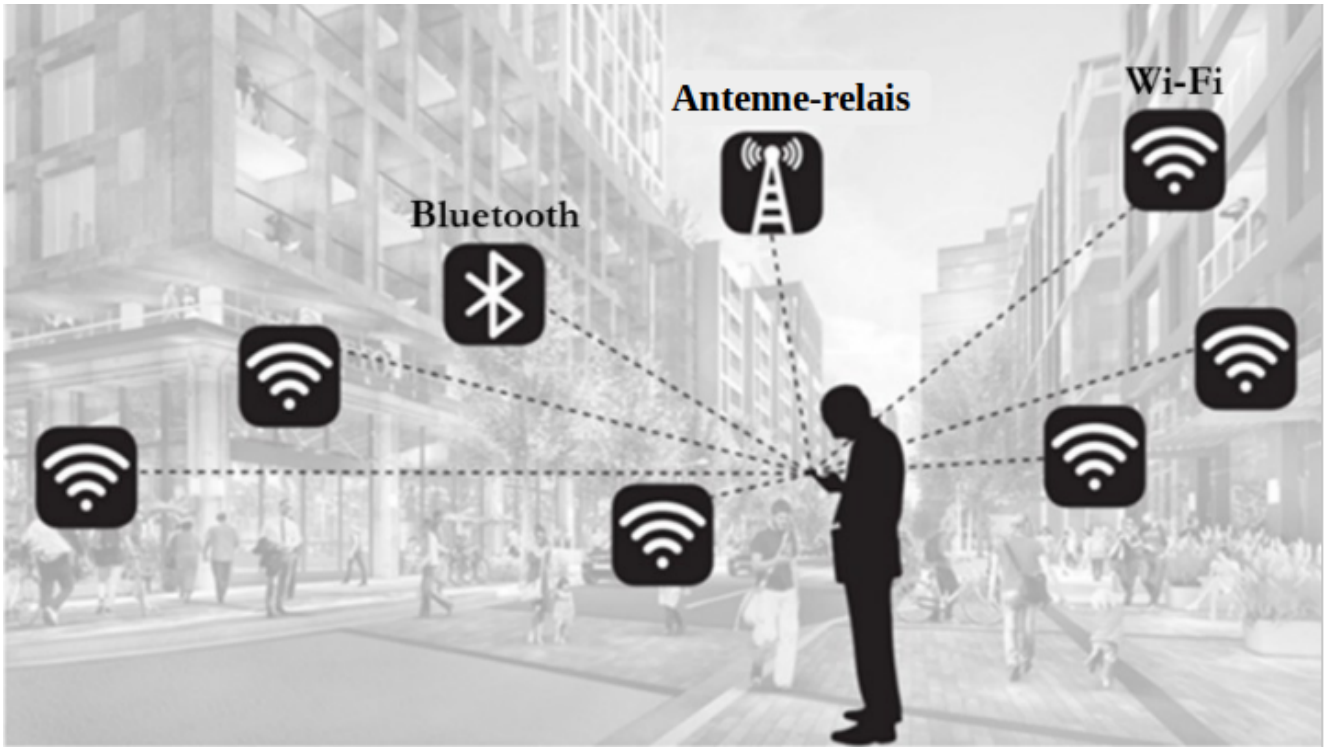

Figure 3 : Android et Chrome utilisent diverses manières de localiser l'utilisateur d'un téléphone.

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.

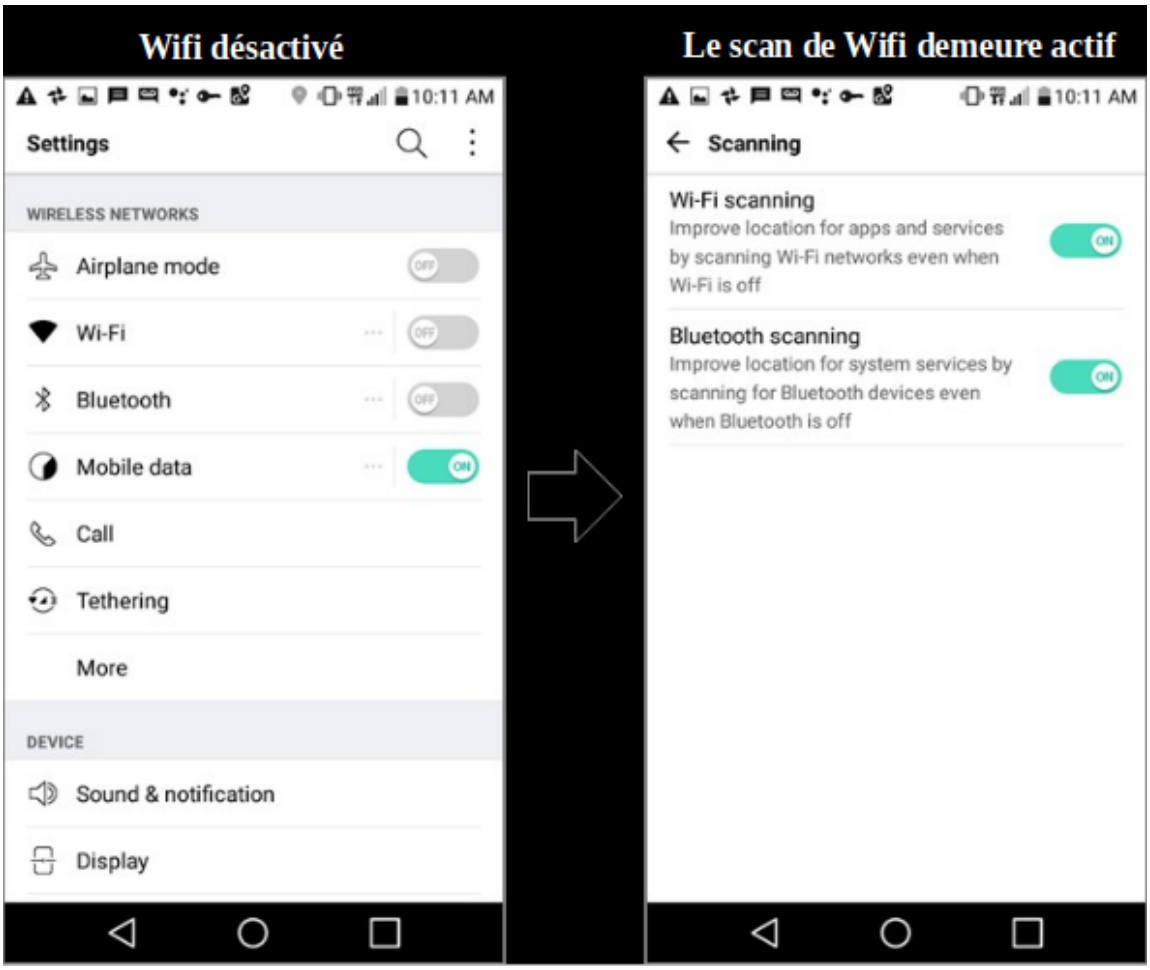

Figure 4 : Android collecte des données même si le Wi-Fi est éteint par l'utilisateur

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 avec les 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.

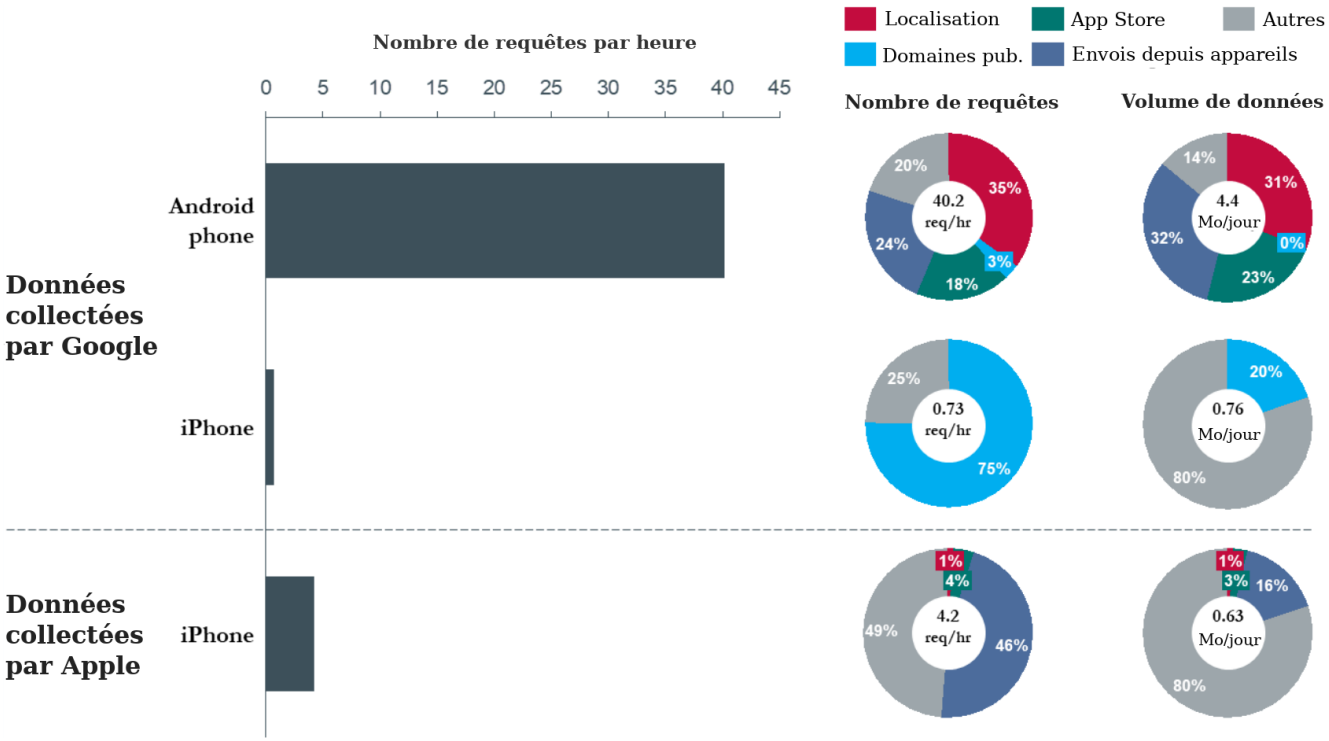

Figure 6 : Données sur le trafic envoyées par les appareils Andoid et les iPhones en veille.

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 aucun 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.