TousAntiCovid : gros plan sur les deux protocoles de traçage numérique de l'application

Vincent Roca

TousAntiCovid : gros plan sur les deux protocoles de traçage numérique de l'application

Vincent Roca

Une application, deux protocoles de traçage numérique

Du traçage manuel de contacts...

Le traçage de contacts est une activité clé pour tenter de contenir les crises épidémiques. À partir d'une personne testée positive à la Covid-19, appelons la Alice, deux types de « contacts » peuvent être recherchés : une personne qu'Alice aurait pu infecter durant la période où elle était contagieuse – c'est le traçage « avant » – ; la personne ayant contaminé Alice – c'est le « rétro-traçage » –, et à partir de là, de nouveau toute personne ayant été en contact avec ce contaminateur.
Or l'analyse des chaînes de contamination fait apparaître que la diffusion se fait essentiellement par un très petit nombre de « super-contaminateurs », responsables de la majorité des contaminations[1]. Identifier rapidement ces super-contaminateurs, parfois asymptomatiques, est donc essentiel pour freiner la propagation, et c'est ce que vise le rétro-traçage.
Traditionnellement menée par des équipes spécialisées *via* des interviews, cette approche manuelle est mécaniquement limitée par les forces humaines en présence, par le temps nécessaire à la collecte des informations, par le caractère non anonyme du procédé qui peut inciter à une autocensure du fait de risques de stigmatisation[2], et par la nécessité de connaitre les personnes croisées pour que l'équipe de traçage puisse les joindre à leur tour.

… jusqu'au traçage numérique

Le protocole ROBERT de traçage de contacts

Afin d'exploiter le potentiel que représente la généralisation des smartphones, notre équipe de recherches PRIVATICS[3], spécialisée dans la protection de la vie privée, a proposé le protocole ROBERT en avril 2020[4]. Ce protocole définit un système numérique de traçage de contacts, entièrement automatique, anonyme pour ses utilisateurs et centré autour de l'Autorité de Santé (ministère des Solidarités et de la Santé) qui a la responsabilité du serveur où sont évalués les risques d'exposition à la Covid-19.
ROBERT repose sur l'échange périodique de petits messages *via* la technologie de communication sans fil Bluetooth[5], utilisée pour connecter un casque sans fil par exemple. Ainsi, les messages émis par le

[1] Dans une étude menée sur deux états Indiens, 71 % des personnes positives n'ont contaminé personne alors que moins de 10 % sont à l'origine de 60 % des contaminations (source : R. Laxminarayan et al., « Epidemiology and transmission dynamics of COVID-19 in two Indian states », Science 370, 691-697, 2020).
[2] Roucaute D. « On est là pour convaincre, pas pour contraindre : le rétro-tracing se déploie pour mieux contrôler l'épidémie de Covid-19. » *Le Monde*, 17 juin 2021.
[3] https://team.inria.fr/privatics/
[4] PRIVATICS team, Inria, France Fraunhofer AISEC, Germany, ROBERT. « ROBust and privacy-presERving proximity Tracing. » Avril 2020 (https://hal.inria.fr/hal-02611265/). Plus d'informations : https://github.com/ROBERT-proximity-tracing/documents
[5] Plus exactement BLE, la déclinaison à faible consommation du Bluetooth, qui est devenue la norme.

smartphone d'Alice contiennent essentiellement un pseudonyme qui est renouvelé tous les quarts d'heures (afin de limiter les risques de suivi).
En parallèle le smartphone d'Alice écoute les messages émis par les autres smartphones, et stocke chacun d'eux en local dans un historique. Si Alice est ultérieurement testée positive et accepte de partager ses informations, elle pourra alors transmettre cet historique de messages reçus au serveur afin que ce dernier évalue le risque de contamination de chaque contact durant la période de contagiosité d'Alice, en tenant compte de la durée et de la proximité estimées de l'exposition.
Ces échanges anonymes remplacent donc totalement les informations fournies aux équipes de traçage manuel par téléphone, notamment l'identité des personnes croisées qui n'a aucun intérêt avec ROBERT : notifier les utilisateurs à risque évite de les identifier. Le bénéfice est clair du point de vue vie privée.

Le protocole CLÉA de traçage de lieux cluster

Le printemps 2021 a vu l'ajout d'une autre forme de traçage, complémentaire, avec le protocole CLÉA[6]. L'approche est totalement différente puisque les communications Bluetooth de ROBERT disparaissent au profit d'une fonction de scan de codes QR, ces codes-barres qui sont affichés à l'entrée des Établissements Recevant du Public (ERP) et qui contiennent essentiellement un pseudonyme lié au lieu.
Avec CLÉA, le smartphone d'Alice mémorise les lieux qu'elle a pu fréquenter, afin de savoir si elle s'est retrouvée, durant la même tranche horaire, co-localisée avec une ou plusieurs personnes testées plus tard positives. Inversement, si Alice est ultérieurement testée positive et accepte de partager ses informations, elle pourra alors transmettre son historique de lieux fréquentés au serveur. Ce dernier pourra trier ces lieux entre ceux qui correspondent à un risque de contagion pour les autres (typiquement 3 jours avant la date de premiers symptômes si elle est connue), et ceux qui correspondent aux lieux où elle a pu être contaminée. Le serveur déduit de ces informations une liste noire de pseudonymes et tranches horaires correspondant aux lieux cluster potentiel, liste qui est publiée et récupérée par chaque smartphone afin d'effectuer une comparaison en local.
Cette approche est certes « moins précise » que ROBERT, puisqu'il n'y a pas d'estimation de distance ou de durée de contact, uniquement une présence dans un lieu qui se révèle à un moment donné cluster. Mais à l'inverse, elle permet à la fois le traçage avant et le rétro-traçage (impossible avec ROBERT), et elle n'est pas limitée par les contraintes technologiques du Bluetooth ce qui permet d'être supporté par un plus grand nombre de téléphones.

Le traçage numérique : un sujet très polémique

Des interrogations légitimes

Proposer un système qui enregistre des contacts (ROBERT) ou des fréquentations de lieux (CLÉA) pour tenter de lutter contre une pandémie, et mettre ce système dans les mains d'un État, ne sont pas des actes neutres, des questions légitimes se posent, et il y a un avant et un après[7]. Cela a suscité de nombreuses réactions et c'est compréhensible. Mais prenons un peu de recul.

Ne rien faire n'était pas envisageable

[6] Roca V., Boutet A., Castelluccia C. « The Cluster Exposure Verification (CLEA) Protocol: Specifications of Protocol Version 0. » Juin 2021 (https://hal-lara.archives-ouvertes.fr/hal-03146022). Plus d'informations : [https://gitlab.inria.fr/stopcovid19/CLEA-exposure-verification]
[7] Par exemple certains sénateurs ont appelé à un usage plus poussé de ces technologies afin de mieux contrôler une éventuelle nouvelle pandémie (https://www.publicsenat.fr/article/societe/covid-19-un-rapport-du-senat-preconise-la-collecte-de-donnees-personnelles-pour).

Devions-nous, avec tous les acteurs du consortium StopCovid/TousAntiCovid[8], proposer une solution dont on a la pleine maîtrise ? Oui, pour plusieurs raisons.
Tout d'abord, ne pas le faire, c'était faire l'impasse sur une technologie qui pouvait aider à limiter la pandémie et sauver des vies, sachant qu'elle est à notre portée et « qu'elle pourrait marcher ». Comment justifier un tel choix ?
La question du rapport bénéfices/risques ne pouvait être tranchée au printemps 2020, il n'existait pas d'antécédent, uniquement des simulations encourageantes conduites par des épidémiologistes[9]. Depuis, une étude portant sur l'application du Royaume-Uni, qui est basée sur GAEN, « Google Apple Exposure Notification »[10], et a bénéficié d'une adoption importante par la population, a montré un impact significatif sur la propagation du virus durant le dernier trimestre 2020 (l'étude estime qu'entre 284 000 et 594 000 cas, suivant la méthodologie utilisée, ont ainsi pu être évités)[11].
Ensuite, ne pas le faire, c'est prendre le risque de subir et devoir accepter les choix faits par d'autres. D'ailleurs, dès le 10 avril 2020, Google et Apple se sont positionnés en publiant les spécifications de la solution GAEN, avant de procéder à une intense activité de lobbying politique, scientifique et technologique, qui a conduit à son adoption partout en Europe, hormis la France. Or la solution GAEN est tout sauf neutre.

À propos de GAEN et de la notion de décentralisation : attention aux intuitions trompeuses

Au cœur de la polémique autour des solutions de traçage de contacts se trouve une « question de spécialistes » : l'architecture doit-elle être centralisée ou décentralisée ? Cette question porte en fait sur la localisation de la fonction d'analyse du risque d'exposition : sur un serveur – approche centralisée, type ROBERT –, ou les smartphones – approche décentralisée, type GAEN ? Et si l'intuition fait pencher vers une décentralisation puisque les informations restent alors en local, cette question amène en fait à deux visions largement opposées.
GAEN repose en effet sur un arbitrage discutable. C'est ce que Google expliquera début octobre 2020, six mois après la sortie des spécifications, *via* G. Hobgen, directeur en charge de la vie privée de Google/Android[12] : « « if we're faced with any kind of trade-off, one of the guiding insights that we've used through this process is that social graph is more sensitive and more privacy risky than infection status data[13]. » Ce choix peut se justifier dans le cas d'États autoritaires. En effet, il est garanti que l'historique de contacts d'Alice, qui se trouve derrière cette notion de « graphe social », restera sur son smartphone, quoi qu'il arrive. C'est le principal bénéfice de l'approche « décentralisée » de GAEN[14].
Malheureusement, la décentralisation nécessite la diffusion des pseudonymes à risque de toutes les personnes qui se sont déclarées positives, afin de pouvoir comparer avec l'historique de contacts en local. Et en pratique, cette liste noire de pseudonymes est librement accessible sur Internet[15]. Or ceci entre en totale contradiction avec le Règlement Général sur la Protection des Données (RGPD) qui reconnaît les données de santé comme étant en haut de l'échelle de protection – ce sont des « données sensibles ».

---

[8] [https://www.inria.fr/fr/stopcovid]
[9] Ferretti *et al*. « Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. » *Science* 368, 619 (2020) 8 May 2020 (https://science.sciencemag.org/content/368/6491/eabb6936).
[10] https://www.google.com/covid19/exposurenotifications/
[11] Wymant C., Ferretti L., Tsallis D. *et al*. »The epidemiological impact of the NHS COVID-19 app » *Nature* 594, 408–412 (2021).
[12] Hobgen G. (Google). « Exposure Notifications: using technology to help public health authorities fight COVID-19. » *Virtual Workshop on Privacy Aspects of Contact Tracing*, October 2, 2020 (https://www.youtube.com/watch?v=0gqZJXOO9Ko) (offset 3:35)
[13] « Si nous sommes confrontés à un compromis, l'une des idées directrices que nous avons utilisées tout au long de ce processus est que le graphe social est plus sensible et plus risqué du point de vue de la vie privée que les données sur le statut positif à la COVID-19. »
[14] Des mauvaises langues pourraient aussi avancer que la décentralisation, c'est l'assurance, pour Google et Apple, de conserver la maîtrise complète du système de traçage et des données collectées, bien loin de l'idée d'une éventuelle neutralité des systèmes d'exploitation.
[15] Kessibi G., Cunche M. *et al*. « Analysis of Diagnosis Key distribution mechanism in contact tracing applications based on Google-Apple Exposure Notification (GAEN) framework." *Document de travail HAL*, septembre 2020 (https://hal.inria.fr/hal-02899412).

Est-ce grave ? Oui, car avec GAEN chacun peut aisément surveiller son entourage, voisins et collègues, et savoir immédiatement si l'un d'eux se déclare positif[16]. Si les risques de discrimination qui en résultent semblent peu préoccupant après deux ans de crise Covid-19 – tout le monde s'est « habitué » au risque –, il n'en a pas été de même au début (certains professionnels en contact avec les malades ont été fortement stigmatisés[17]), et cela pourrait changer si la dangerosité du virus devait augmenter significativement. Négliger la protection des données de santé n'est pas neutre, et cette décision n'appartient en aucun cas à une société privée, *a fortiori* étrangère.
Enfin placer toute sa confiance dans les mains d'un acteur tiers n'est une garantie absolue. Ainsi Google a laissé fuiter les données clés de GAEN – en l'occurrence tous les pseudonymes transmis et reçus ainsi que le statut de santé de l'utilisateur – pendant près d'un an, dans les traces systèmes de tous les téléphones Android, accessibles à plusieurs centaines d'applications préinstallées[18].

À la recherche d'un compromis équilibré avec TousAntiCovid

Dans ce contexte, TousAntiCovid nous semble plus que jamais une solution raisonnable, qui propose un compromis équilibré et assumé.
Ce qui a été fait au printemps 2020 l'a été dans un contexte précis, la France de 2020, et avec un objectif précis, aider un État démocratique à faire face à cette situation inédite. ROBERT n'a jamais été pensé comme une solution universelle car trois hypothèses sont faites :

- il existe des institutions de confiance qui posent un cadre légal démocratiquement discuté – ainsi des discussions parlementaires ont eu lieu préalablement à tout déploiement en France, en mai 2020 – ;
- il existe une autorité de protection de données indépendante, la CNIL, en capacité de conduire des audits impartiaux et de s'opposer à d'éventuelles dérives ;
- enfin, l'usage de l'application et des fonctions de traçage est volontaire : la confiance des utilisateurs se mérite et toute dérive est sanctionnée.

Fort de ces hypothèses, ROBERT, en étant centralisé, protège par conception à la fois le graphe social et le statut de positivité des utilisateurs qui se déclarent, car :

- les informations d'historique remontées au serveur sont minimales et mélangées avec les autres remontées, afin de rendre impossible toute velléité de reconstruction de graphe social ;
- les informations de santé sont protégées par le serveur central, hors de portée des attaquants.

La centralisation, dans un État démocratique et sous le contrôle d'une autorité indépendante de protection des données, permet nativement à la fonction de traçage de contacts d'atteindre cette double protection, graphe social et données de santé.
À l'inverse de ROBERT, le protocole CLÉA repose sur une approche décentralisée. La raison est simple : les données diffusées publiquement pour permettre cette évaluation du risque sur le smartphone, à savoir la liste noire des pseudonymes des lieux cluster, ne sont pas des données sensibles, de santé, car elles ne sont pas attachées à des personnes physiques. C'est une différence fondamentale !

Pour finir, à propos de l'importance d'une solution souveraine

TousAntiCovid n'est pas une solution par défaut d'un État qui s'oriente vers une « solution sur étagère ». Au contraire, TousAntiCovid repose sur un triptyque : souveraineté, efficacité, et respect de la vie privée. Elle a permis de donner toute son importance à l'autorité de santé, placée au centre. Le ministère des Solidarités et de la Santé est ainsi le responsable de traitement, maîtrisant des choix technologiques et des données collectées, afin de s'adapter au mieux à la pandémie.

---

[16] Un simple navigateur web suffit, à la fois pour collecter les petits messages Bluetooth puis pour les comparer avec la liste noire librement accessible sur Internet. Voir le démonstrateur coronadetective.eu (https://www.coronadetective.eu/).
[17] Léo F. « Seine-et-Marne : une infirmière priée de quitter son domicile par un voisin. » *Le Parisien*, 26 mars 2020.
]

C’est une liberté essentielle quand on pense qu’au Royaume-Uni, une mise à jour de l’application nationale qui visait à ajouter la fonction de traçage de lieux cluster du type CLÉA a été refusée par Google et Apple car contraire aux règles que ces sociétés ont imposées en matière de collecte de données de localisation[19].
La solution est-elle parfaite ? Certainement pas, elle résulte de compromis, parfois en arbitrant entre des contraintes opposées. Un exemple : avoir une solution anonyme, pour protéger les utilisateurs, n’est pas négociable. Mais cet anonymat ouvre aussi la porte à certaines attaques qui pourraient sinon être limitées (on peut penser à un « marché noir » des preuves de positivités, anonymes, présentées au serveur quand un utilisateur partage son historique de contacts ou de lieux visités). Des compromis de ce type sont inévitables et ils sont assumés.
Pour conclure, Shoshana Zuboff, Professeure émérite à la Harvard Business School, à l’origine de la théorisation du « capitalisme de surveillance », a expliqué: « En réalité, les deux approches [centralisée et décentralisée] ont leurs atouts et leurs inconvénients. Les deux ! Mais la différence est qu'une de ces approches est sous le coup de la loi et de l'État de droit. L'autre approche est celle de sociétés privées non régulées dont la fortune vient du capitalisme de surveillance[20]. »
Finalement, ce qu’a permis TousAntiCovid, c'est cette liberté de choix, essentielle pour face à une crise inédite, dans le respect de nos valeurs.

[19] Kelion L. « NHS Covid-19 app update blocked for breaking Apple and Google's rules. » *BBC New*s, 12 avril 2021.
[20] Zuboff S. « Il y a dans l'ombre une énorme extraction de nos données personnelles dont nous n'avons pas idée. » *Le Figaro*, octobre 2020.