Deux ou trois choses sur les applications de suivi de contacts pendant l'épidémie

Notre petit dossier sur l'application de contact-tracing « StopCovid » s'enrichit aujourd'hui d'un article de Stéphane Bortzmeyer, que nous avons déjà invité à de multiples reprises sur le Framablog.

La position de Framasoft au sujet de cette application est plutôt claire : StopCovid est un leurre de communication politique.

Cependant, nous trouvons important d'apporter au débat des articles plus pédagogiques. Ce que nous avons fait par exemple en publiant la bande dessinée de Nicky Case, et aujourd'hui avec cet article qui, nous l'espérons vous éclairera un peu plus dans le débat.

⇒ Accéder aux articles déjà publiés dans notre dossier StopCovid

(NB : cet article est une republication, avec l'accord de son auteur, de l'article https://www.bortzmeyer.org/tracking-covid-19.html dont la première publication est datée du 19 avril 2020 — Nous vous encourageons à vérifier sur le site originel si l'article a pu être mis-à-jour depuis).

Deux ou trois personnes m'ayant demandé si j'avais une opinion sur les **applications de suivi de contacts**, dans le contexte de **l'épidémie de COVID-19**, je publie ici quelques notes, et pas mal d'hyperliens, pour vous donner de la lecture pendant le **confinement**.

D’abord, des avertissements :

- Je ne suis pas **épidémiologiste**. Même si je l’étais, il y a encore beaucoup de choses que la science ignore au sujet des infections par le **SARS-CoV2**, comme les durées exactes des phases où on est contagieux.
- Je ne suis pas non plus un spécialiste de la conception et de l’analyse de **protocoles** de suivi de contacts. Mais ce n’est pas très grave, pour les raisons que j’exposerai rapidement par la suite.
- Vous ne trouverez pas ici d’analyse de l’application annoncée par le gouvernement français, StopCovid, pour la bonne et simple raison qu’elle n’existe pas. Il y a eu des promesses sous la forme de quelques mots (« anonyme », « sur la base du volontariat ») mais aucun détail technique n’a été publié. À l’heure actuelle, il n’est donc pas possible de dire quoi que ce soit de sérieux sur cette application spécifique.
- D’une manière générale, la situation évolue vite, et il est très possible que, dans une ou deux semaines, cet article ne vaille plus rien.

Maintenant, rentrons dans le vif du sujet. Dans la description et l’analyse des protocoles comme PACT, DP3T ou ROBERT ? Non, car, pour moi, c’est une question très secondaire. Voyons les problèmes par ordre décroissant d’importance.

D’abord, il faut se demander si une telle **application de suivi des contacts** est utile et, surtout, si elle justifie les efforts qu’on y consacre, par rapport à des sujets moins *high-tech*, moins prestigieux, moins *startup-nation*, qui motivent moins les informaticiens mais qui ont plus d’importance pour la santé publique, comme la production et la distribution de **masques**, ou comme la revalorisation des rémunérations et des conditions de travail du personnel de santé. Je ne vais pas insister sur ce point, c’est certes le plus important mais **la Quadrature du Net** en a déjà parlé, et mieux que moi.

Bref, ce projet d'application de suivi des contacts semble davantage motivé par le désir d'agir, de faire quelque chose, même inutile, désir qui est commun en temps de crise, plutôt que par un vrai problème à résoudre. (C'est ce que les anglophones nomment la maladie du *do-something-itis*.) Il y a également des enjeux commerciaux, qui expliquent que certaines entreprises se font de la publicité en affirmant travailler sur le sujet (sans tenir compte des travaux existants).

Mais surtout, une application n'a de sens que si on teste les gens, pour savoir qui est contaminé. Comme on peut apparemment être contagieux et pourtant asymptomatique (pas de maladie visible), il faut tester ces personnes asymptomatiques (qui sont sans doute celles qui risquent de contaminer le plus de gens puisque, ignorantes de leur état, elles sortent). Or, **Macron** a bien précisé dans son discours du 13 avril qu'on ne testerait pas les personnes asymptomatiques (probablement car il n'y a pas de tests disponibles). Cela suffit à rendre inutile **toute** application, indépendamment des techniques astucieuses qu'elle utilise, car l'application elle-même ne peut pas déterminer qui est malade ou contagieux.

Ensuite, le protocole est une chose, la mise en œuvre dans une application réelle en est une autre. Le diable est dans les détails. Comme indiqué plus haut, on ne sait encore rien sur l'application officielle, à part son nom, StopCovid. Pour formuler un avis intelligent, il ne faudra pas se contenter de généralités, il faudra regarder son code, les détails, les traqueurs embarqués (une plaie classique des applications sur **ordiphone**, cf. le projet ExodusPrivacy), etc. Il faudra aussi se pencher sur le rôle du **système d'exploitation** (surtout s'il y a utilisation de l'**API** proposée par Google et Apple). Le fait que l'application soit en **logiciel libre** est évidemment un impératif, mais ce n'est pas suffisant.

Si vous n'êtes pas informaticienne ou informaticien, mais que vous voulez vous renseigner sur les applications de suivi de contacts et ce qu'il y a derrière, souvenez-vous qu'il y a

plusieurs composants, chacun devant être étudié :

- L'application elle-même, celle que vous téléchargez sur le **magasin**, qui est la partie visible (mais pas forcément la plus importante).
- Le **protocole** qui est l'ensemble des règles que suit l'application, notamment dans la communication avec le reste du monde (autres **ordiphones**, serveur central…). Avec le même protocole, on peut créer plusieurs applications assez différentes.
- Le **système d'exploitation** qui, après tout, a un complet contrôle de la machine et peut passer outre les décisions des applications. C'est un sujet d'autant plus sensible que, sur les ordiphones, ce système est étroitement contrôlé par deux entreprises à but lucratif, Apple et Google.
- Le serveur central (la grande majorité des protocoles proposés nécessite un tel serveur) qui peut être piraté ou, tout simplement, géré par des gens qui ne tiennent pas leurs promesses.

Parmi les bonnes lectures accessibles à un large public :

- Un article de Martin Untersinger.
- Le texte de Paula Forteza et Elliot Alderson.

Voilà, on peut maintenant passer aux questions qui passionnent mes lecteurs et lectrices passionnés d'informatique, les protocoles eux-mêmes. Il en existe de nombreux. J'ai une préférence pour PACT, dont je vous recommande la lecture de la spécification, très claire. La proposition DP3T est très proche (lisez donc son livre blanc).

Ces deux propositions sont très proches : l'ordiphone émet en **Bluetooth** des identifiants temporaires, générés aléatoirement et non reliables entre eux. Les autres ordiphones proches les captent et les stockent. Ces identifiants se nomment *chirps* dans PACT (qu'on pourrait traduire par « cui-cui ») et EphID

(pour *Ephemeral ID*) dans DP3T. Lorsqu'on est testé (rappel : il n'y a pas assez de tests en France, on ne peut même pas tester tous les malades, ce qui est un problème bien plus grave que le fait d'utiliser tel algorithme ou pas), et détecté contaminé, on envoie les données à un serveur central, qui distribue la liste. En téléchargeant et en examinant cette liste, on peut savoir si on a été proche de gens contaminés.

C'est évidemment une présentation très sommaire, il y a plein de détails à traiter, et je vous recommande de ne pas vous lancer dans de longues discussions sur **Twitter** au sujet de ces protocoles, avant d'avoir lu les spécifications complètes. Les deux propositions ont été soigneusement pensées par des gens compétents et le Café du Commerce devrait lire avant de commenter.

PACT et DP3T ont assez peu de différences. Les principales portent sur le mécanisme de génération des identifiants, PACT déduit une série d'identifiants d'une **graine** renouvelée aléatoirement (on stocke les graines, pas réellement les identifiants), alors que DP3T déduit chaque graine de la précédente, des choses comme ça.

La proposition ROBERT est assez différente. La liste des identifiants des contaminés n'est plus publique, elle est gardée par le serveur central, que les applications doivent interroger. Globalement, le serveur central a bien plus de pouvoir et de connaissances, dans ROBERT. La question est souvent discutée de manière binaire, avec centralisé vs. décentralisé mais le choix est en fait plus compliqué que cela. (Paradoxalement, un protocole complètement décentralisé pourrait être moins bon pour la vie privée.) Au passage, j'ai déjà discuté de cette utilisation très chargée de termes comme « centralisé » dans un article à JRES. Autre avantage de ROBERT, la discussion sur le protocole se déroule au grand jour, via les tickets de **GitHub** (cf. leur liste mais lisez bien la spécification avant de commenter, pas juste les images). Par contre, son analyse de sécurité est très

insuffisante, comme le balayage de tous les problèmes liés au serveur central en affirmant qu'il sera « honnête et sécurisé ». Et puis la communication autour de cette proposition est parfois **scientiste** (« Ce sont des analyses scientifiques qui permettent de le démontrer, pas des considérations idéologiques [comme si c'était mal d'avoir des idées, et des idées différentes des autres] ou des a priori sémantiques. ») et il y a une tendance à l'exagération dans les promesses.

Enfin, un peu en vrac :

- Le Wikipédia anglophone a **un bon guide des applications de suivi**.
- La proposition **Apple/Google**, détaillé ici ainsi que sur cette page. Il ne s'agit à proprement parler pas d'un protocole de suivi de contacts mais d'un ensemble de services dans les **systèmes d'exploitation** (**iOS** pour Apple et **Android** pour Google), pour aider les applications de suivi.
- Je n'ai pas cité les autres protocoles, NTK et TCN.
- La question de la protection de la **vie privée**, avec de telles applications, a suscité bien des discussions (et à juste titre). L'**Union européenne** a produit une série de recommandations à ce sujet, et le **CCC** une autre série d'exigences à satisfaire.
- La mise en correspondance de deux ensembles de données (les identifiants des gens avec qui il y a eu contact, et les identifiants des gens contaminés) est un problème très difficile à résoudre si on veut respecter un peu la vie privée. Mais il n'est pas insoluble, voir par exemple la solution de Signal.
- Un bon article de **Ross Anderson**, qui prend de la hauteur sur le sujet.