« Contact tracing » : quelques éléments pour mieux comprendre les enjeux

Bruno Sportisse, PDG d'Inria

Le numérique est pleinement mobilisé dans la lutte contre le Covid19 : chercheurs, développeurs, entreprises, se sont, dès les premiers jours, engagés aux côtés des personnels soignants et des médecins qui travaillent sur des solutions médicales (vaccins et traitements), pour leur apporter leurs compétences et leurs expertises. Au sein d'Inria, l'institut national de recherche pour les sciences et technologies du numérique, que j'ai l'honneur de diriger, plus d'une vingtaine de projets, avec nos partenaires de la recherche publique (CNRS, CEA, INSERM, INRAE, Universités, Ecoles d'ingénieurs) et souvent des entreprises, ont ainsi été initiés, avec des composantes souvent très opérationnelles, pour aider les hôpitaux dans la gestion numérique de la crise ou accompagner la recherche médicale.

La mobilisation spontanée des scientifiques et des ingénieurs du numérique a été exceptionnelle. Cette prise de parole est aussi l'occasion de leur rendre hommage.

Parmi les projets, la question du « traçage numérique » (« contact tracing ») est à présent au cœur des préoccupations, depuis que les Ministres en charge respectivement de la Santé et du Numérique, Messieurs Olivier Véran et Cédric O, ont annoncé le 8 avril 2020 qu'un travail était mené pour construire le prototype d'une application française, STOPCOVID, dans le cadre d'une stratégie globale de déconfinement. Le leadership du projet, qui associe acteurs publics et privés, a été confié à Inria.

La France participe, au travers d'Inria, à l'initiative PEPP-PT aux côtés d'équipes allemandes, italiennes et suisses. Les équipes d'Inria publient aujourd'hui, conjointement avec nos partenaires du Fraunhofer, le protocole ROBERT - pour ROBust and privacy-presERving proximity Tracing – qui représente l'état de l'art de nos réflexions sur l'architecture technique d'une application de « contact tracing » respectueuse des valeurs européennes. Ce protocole est disponible sous Github (https://github.com/ROBERT-proximity-tracing/) comme les pratiques scientifiques standard le veulent.

Ce texte a vocation à expliquer en des termes compréhensibles par tous ce que contient ce protocole, mais surtout les présupposés et l'esprit dans lequel il a été conçu. Il est important, dans l'urgence qui caractérise les circonstances exceptionnelles que nous vivons, de faire le point, sereinement, sur un sujet difficile, qui brasse des dimensions multiples.

Afin de poser le cadre, il me semble utile de commencer par rappeler ce qu'une application qui reposerait sur ce protocole n'est pas, eu égard aux interrogations légitimes qui s'expriment et aux confusions qui peuvent avoir lieu.

Une telle application n'est pas une application de « tracking » : elle n'utilise que le bluetooth, en aucun cas les données de bornage GSM ni de géolocalisation.

Une telle application n'est pas une application de surveillance : elle est totalement anonyme. Pour être encore plus clair : sa conception permet que PERSONNE, pas même l'Etat, n'ait accès à la liste des personnes diagnostiquées positives ou à la liste des interactions sociales.

entre les personnes. La seule information qui m'est notifiée est que mon smartphone s'est trouvé dans les jours précédents à proximité du smartphone d'au moins une personne qui a, depuis, été testée positive et s'est déclarée dans l'application.

Une telle application n'est pas une application de délation : dans le cas où je suis notifié, je ne sais pas qui est à l'origine de la notification. Lorsque c'est moi qui me déclare positif, je ne sais pas qui est notifié.

Une telle application n'est pas obligatoire. Ses utilisateurs choisissent de l'installer. Ils choisissent d'activer le bluetooth. Ils peuvent, à tout moment, désactiver le bluetooth ou désinstaller l'application.

Ces éléments étant précisés, de quoi parle-t-on ?

Par « contact tracing », on désigne la capacité à pouvoir informer une personne, à travers une application présente sur son smartphone, qu'elle a été au contact lors des jours précédents (typiquement de deux à trois semaines) de personnes qui ont été diagnostiquées positives au Covid19. Ce « cas contact » présente, de ce fait, un risque d'être porteur du virus et d'accélérer la diffusion de l'épidémie. Les « moyens numériques » qui permettent de qualifier ce risque reposent sur la capacité de deux smartphones à reconnaitre qu'ils sont à proximité l'un de l'autre, à travers la technologie bluetooth, qui n'est opérante qu'à faible distance (quelques mètres). De nombreux projets préfèrent ainsi parler de « proximity tracing », plus précis sur le rôle joué par les smartphones, terme que j'adopterai dans la suite. Aucune technologie de géolocalisation (à tel lieu, à telle heure) n'est ainsi mise en œuvre.

Une fois défini ainsi, le « proximity tracing » pose évidemment de nombreuses questions sur son usage : comment une personne diagnostiquée positive utilise-t-elle cette information ? quelles sont les consignes à tenir suite à cette information par/pour les personnes identifiées à risque dans le cadre d'une stratégie globale de santé ? Nous reviendrons sur ces points plus loin.

A quoi est-ce que cela peut servir ?

Les services médicaux ont une longue pratique des enquêtes de terrain pour retrouver les chaines de propagation d'une épidémie. En tant que tel, le type d'application visé n'est donc à voir que comme une aide complémentaire à ces pratiques. Un exemple significatif de déploiement est celui de l'application Trace Together à Singapour.

Plus récemment, des travaux scientifiques, menés notamment par un épidémiologiste d'Oxford, Christophe Fraser, ont montré, *sur la base de simulations*, que l'utilisation d'une utilisation de « proximity tracing » était une aide utile pour casser la propagation de l'épidémie. Son équipe, multidisciplinaire, a ainsi simulé l'évolution pendant 250 jours d'une ville fictive (modélisée) d'un million d'habitants et a montré l'impact d'une utilisation de l'application en fonction de plusieurs niveaux de diffusion de l'application (de 0 à 80%). Pour résumer très simplement les résultats de cette simulation, le téléchargement de l'application par une ou deux personnes (selon les cas) entraine la réduction de la transmission du virus à une personne. Bien sûr, de nombreux paramètres sont susceptibles d'influer l'impact (comme

par exemple les « faux positifs » de transmission ou encore la nature du respect des gestes barrières). Ces travaux montrent une tendance, qui est conforme au bon sens.

En tout état de cause, aucune équipe travaillant sur ces sujets n'oublie que le « proximity tracing » n'est qu'une composante d'un ensemble plus vaste de mesures, dans le cadre d'une approche pilotée par une politique de santé. Je ne connais personne qui croie au solutionnisme technologique en la matière.

Quelles sont les composantes d'un tel système ?

A cette humilité de l'approche, il convient d'ajouter, en transparence, la présentation de l'ensemble des dimensions d'un tel système.

Pour commencer, il y a des **limitations technologiques** : la technologie bluetooth n'a pas été conçue pour être précise dans l'estimation de distances entre deux smartphones. Les résultats peuvent dépendre de nombreux paramètres, comme la physiologie des personnes, la position du smartphone, le type de smartphone, l'état de la batterie, etc. Cela a conduit plusieurs équipes internationales à mener des tests de calibration pour proposer des modèles statistiques qui corrigent ces erreurs. C'est par exemple le cas des équipes allemandes dans le cadre de l'initiative européenne PEPP-PT (sur laquelle je reviendrai).

Une autre limitation est liée au **modèle de transmission du virus**, qui reste très incertain : via des aérosols ou des gouttelettes (plus grosses), avec un impact sur le temps de résidence dans l'air ? via des surfaces ? comment estimer la charge virale ? etc. Toutes les applications de « proximity tracing » reposent ainsi sur des fonctions de risque, définies, avec les chercheurs en épidémiologie, sur la base de l'état de l'art. Cette connaissance est encore très lacunaire et est susceptible de changer très rapidement, en fonction des retours d'expérience.

Enfin, un sujet clé et extrêmement sensible est celui du **protocole d'échanges d'informations**. C'est un sujet extrêmement sensible car il touche à des dimensions politiques et démocratiques : quelles informations sont susceptibles d'être transmises et à qui ? Quelles sont les autorités/les organisations qui opèrent ces systèmes de transmission ? A qui fait-on confiance ? Quels sont les niveaux de sécurité et les hypothèses d'attaques que nous sommes prêts à envisager et à assumer de manière crédible ?

Premier point pour commencer : aucun projet n'a pour ambition de mettre en place un réseau de pair-à-pair, où tout reposerait sur une communauté supposée « indépendante » (je reviendrai sur ce point) de terminaux/de smartphones qui échangent des informations entre eux. La raison principale est l'impact des failles de sécurité qui pourraient exister avec une telle approche.

Tous les systèmes projetés comportent donc une composante commune (un serveur) et une composante décentralisée (un ensemble de smartphones qui peuvent communiquer entre eux à travers le bluetooth) : **tous les systèmes actuellement étudiés sont donc à la fois centralisés** (je reviens ci-après sur l'utilisation du terme) **et décentralisés**.

Dans ce contexte, les débats sur les avantages supposés d'un système *parce qu'il serait décentralisé* vis-à-vis d'un autre système *parce qu'il serait centralisé* ne me semblent pas relever du champ de la rigueur scientifique. Le terme « centralisé » est souvent utilisé à dessein, en stigmatisant implicitement un Etat supposé vouloir être traqueur. Des approches supposées être très décentralisées, qui pourraient avoir les faveurs de communautés réticentes à accorder leur confiance à une autorité centrale, peuvent présenter des faiblesses majeures en matière de protection de la vie privée. Ce sont des analyses scientifiques, par définition vérifiables et se prêtant à une discussion, qui permettent de le démontrer, pas des considérations idéologiques ou des a priori sémantiques.

Deuxième point, quelles informations circulent ?

Dans tous les projets respectueux du cadre européen de protection de la vie privée (le RGPD, sous le contrôle des autorités indépendantes comme la CNIL en France), les informations circulent sous la forme de « crypto-identifiants », des données pseudonymisées, en général générées de manière éphémères (typiquement, pour une période de 15 minutes) et associées à un terminal *et non à une personne*. C'est-à-dire qu'un smartphone va rencontrer au cours de ses déplacements des crypto-identifiants éphémères (ceux des smartphones rencontrés).

Les projets existants diffèrent sur la nature des informations transmises à un serveur central, opéré par une autorité de santé, un tiers de confiance voire une entreprise.

Ce jour, les équipes françaises et allemandes d'Inria et du Fraunhofer viennent ainsi, comme évoqué plus haut, de sortir un protocole conjoint, dénommé ROBERT pour ROBust and privacy-presERving proximity Tracing.

Je veux prendre quelques lignes pour citer les collègues d'Inria, avec lesquels j'ai l'honneur de travailler depuis 3 semaines, aux côtés d'Eric Fleury, embarqué aussi dans cette aventure : Nataliia Bielova, Antoine Boutet, Claude Castelluccia, Mathieu Cunche, Cédric Lauradoux, Daniel Le Métayer et Vincent Roca (par ordre alphabétique). Ils sont l'honneur de la recherche française en informatique, en manipulant à la fois des concepts informatiques de haut niveau (liés aux protocoles cryptographiques et aux protocoles de transmission) et une compréhension fine des enjeux éthiques et sociétaux. Le nom de leur équipe, PRIVATICS, se suffit à lui-même. Je ne veux pas oublier non plus les collègues allemands, Claudia Eckert, Alexander Küchler, Martin Schanzenbach et Julian Schütte de AISEC, l'institut de cybercurité de la Fraunhofer Gesellschaft. C'est aussi une belle illustration du partenariat franco-allemand.

Ce protocole repose sur les principes suivants, ce qui me permet de rentrer dans le détail sur un exemple :

- Une personne, soucieuse de participer à la lutte contre la propagation de l'épidémie, télécharge *de manière volontaire* l'application sur son smartphone. Son smartphone reçoit alors un ensemble de crypto-identifiants (ou une méthode pour les générer toutes les 15 minutes).
- Le détenteur du smartphone, en laissant le bluethooth activé, permet à son application de construire un historique des crypto-identifiants rencontrés, « à proximité », pendant une durée significative lors des déplacements (ces crypto-

identifiants étant sur les smartphones des personnes ayant elles-aussi téléchargé l'application).

- Si la personne est diagnostiquée positive, elle fait remonter son historique de crypto-identifiants rencontrés sur un serveur d'une autorité de santé (par exemple), sans divulguer au serveur son (ses) propre(s) crypto-identifiant(s). Aucun lien n'est fait entre le téléphone de la personne et son historique. Chacun de ces crypto-identifiants est donc potentiellement « à risque » (il correspond, sans qu'aucun lien ne soit possible avec une personne, à un smartphone qui a été en proximité d'un smartphone porté par une personne qui a été ultérieurement diagnostiquée positive).
- Par ailleurs, chaque smartphone ayant téléchargé l'application vérifie auprès du serveur central, « de temps à temps » (toutes les heures, tous les jours, cela fait partie des paramètres à fixer) si ses propres crypto-identifiants sont parmi ceux à risque. Si c'est le cas, cela signifie que le smartphone a été à proximité lors des jours précédent d'un smartphone porté par une personne qui s'est avérée être positive ultérieurement.
- La notification se fait sur la base d'une évaluation du risque (dont le calcul doit être défini avec les épidémiologistes, je l'ai déjà évoqué) en utilisant l'information de proximité. La flexibilité du système est clé pour le management d'une crise sanitaire, la prise en compte de connaissance médicales qui évoluent rapidement, voire un apprentissage par le système (toujours sur la base de données statistiques anonymisées) pour le rendre plus efficace, par exemple, pour diminuer l'occurrence de faux positifs. Pour autant que l'autorité sanitaire ait la main sur tout cela (je reviens sur ce point-clé ensuite).
- Cette information (c'en est une) peut alors déclencher un renvoi vers divers actes (ce n'est pas le sujet de cet article) : un respect scrupuleux des gestes barrières, un suivi journalier des symptômes, une consultation, un test, etc. Ceci relève du choix d'une politique de santé d'un Etat.

Dans cette approche, plusieurs choix forts ont été effectués, sur lesquels je veux revenir, car ces travaux ont été menés dans un contexte certes sous pression mais dans un cadre de valeurs : tout repose sur le volontariat et le consentement ; on ne doit pas pouvoir inférer que mon voisin/ma voisine est positif/positive voire m'aurait contaminé ; à l'occasion du déploiement d'une telle application, un serveur ne doit pas collecter la liste des personnes contaminées (ce point est vraiment important); le choix d'une politique de santé relève du choix d'un Etat souverain.

Par exemple, sur le serveur central (pour assumer le terme), il n'y a AUCUNE donnée relative au statut des personnes positives. Il s'y trouve une liste de crypto-identifiants des smartphones s'étant trouvés à proximité des smartphones des personnes positives.

Autre exemple de choix fort : dans le smartphone de mon voisin, il n'y a AUCUNE donnée concernant mon diagnostic médical, aussi encrypté soit-il. Il y a une liste des crypto-identifiants de tous les smartphones rencontrés.

Autre exemple relevant de la maitrise d'une politique de santé : les paramètres du modèle de transmission et les données statistiques anonymes sont entre les mains de l'autorité de santé qui fixe l'utilisation de ce système. Pas d'une compagnie privée, aussi innovante soit-elle.

D'autres choix sont possibles et fondent d'autres approches, qui n'ont pas été celles des équipes d'Inria et du Fraunhofer : des crypto-identifiants des personnes diagnostiquées positives peuvent transiter par des serveurs centraux et être envoyés dans tous les smartphones. Smartphones au sein desquels la mise en correspondance entre crypto-identifiants rencontrés et crypto-identifiants de personnes testées positives peuvent être effectués. C'est un système que l'on peut présenter comme fortement décentralisé… tout comme on peut le présenter comme une centralisation décentralisée : il y aura ainsi, sur chaque smartphone, la liste de l'ensemble des crypto-identifiants des personnes diagnostiquées comme positives. Ce système a par ailleurs l'avantage d'être facilement permis par l'API à venir (mi-mai), dévoilée par Apple et Google il y a une semaine, une grande première dans l'histoire de l'informatique.

En tout état de cause, c'est le choix d'un Etat de décider d'utiliser ou non le protocole qu'il désire en fonction de sa politique. Et c'est notre responsabilité de scientifique de lui procurer les moyens de ce choix.

Ce protocole est-il définitif et totalement finalisé ?

La réponse est : non. C'est un travail en cours, qui a vocation à être challengé.

D'abord, comme tout projet scientifique, ce protocole va être soumis à la critique de ses pairs. Cela nécessite une démarche d'ouverture : l'article scientifique est mis à disposition de la communauté scientifique sous Github (https://github.com/ROBERT-proximity-tracing/). Des failles, des attaques, des suggestions vont être proposées. Beaucoup a déjà été fait, notamment à travers les échanges avec l'ANSSI, dont je tiens à saluer l'engagement remarquable.

Ensuite, parce que, comme certains l'ont noté, les applications de « contact tracing » posent un certain nombre de prérequis qui ne peuvent se déployer en l'état sur tous les smartphones utilisés par les Français. Notamment pour que le bluetooth puisse fonctionner de manière efficace même quand le smartphone est en veille. Nous avons donc un besoin impératif de travailler, en bonne intelligence, avec les concepteurs de systèmes d'exploitation (OS) pour ouvrir cette possibilité.

Des nouvelles versions sont donc à venir mais, en tout état de cause, une première implémentation logicielle est en cours de développement sur la base du protocole ROBERT. Comme toutes les productions associées au projet STOPCOVID, elle sera mise en open source, sous licence MPL. De par son histoire, Inria sait combien le logiciel libre permet à une technologie logicielle d'être améliorée et d'être transparente. D'être reprise par d'autres pays, aussi, qui n'auraient pas la capacité à développer de tels logiciels. C'est aussi tout le combat des standards ouverts du WEB, à travers le W3C dont j'ai l'honneur de présider le nœud européen, qui illustre l'attachement d'Inria à ces valeurs.

Un autre enjeu est celui de l'interopérabilité.

C’est tout le sens de la participation de la France, au travers d’Inria, à l’initiative PEPP-PT, déjà évoquée, aux côtés d’équipes allemandes, italiennes et suisses. Nous ne sommes pas toujours d’accord sur les choix effectués, sur les hypothèses effectuées (par exemple : qui est le plus susceptible de mener une attaque ? Un Etat démocratique ou un hacker, même pas très dégourdi, en tout cas pas toujours éthique). Nous n’avons pas nécessairement la même attention aux questions de souveraineté technologique et numérique. Peu importe, nous travaillons ensemble, sur un terrain de jeu scientifique et technologique, ensemble, pour construire des solutions respectueuses des valeurs que nous partageons et inter-opérables (les applications déployées auront des briques communes mais seront nationales du fait de l’importance de l’inscription dans un système de santé national).

En ces moments exceptionnels pour notre Nation, la France peut compter sur son écosystème de recherche et d’innovation pour mener à bien des projets au croisement de la nécessaire efficacité des politiques de santé, du respect des libertés individuelles et du maintien voire du renforcement de notre souveraineté technologique et numérique.