Proposition d'un protocole de suivi des contacts rapprochés dans la lutte contre le COVID-19.

Cette proposition est inspirée du protocole ROBERT disponible ici :
https://github.com/ROBERT-proximity-tracing
ROBERT-summary-FR.pdf
Nous vous en conseillons la lecture.

Nous avons parcouru les commentaires qui ont été apportés depuis sa publication par l'INRIA le 18 avril (en ne comprenant pas forcément toutes les subtilités, certaines interventions étant l'œuvre de spécialistes en sécurité et cryptologie) :
https://github.com/ROBERT-proximity-tracing/documents/issues/6

Pour mettre à profit ce temps de confinement et parce-que l'exercice intellectuel nous paraissait intéressant, nous avons tenté d'élaborer une solution qui nous semble améliorer le système proposé par l'INRIA.

A la lecture de ce document : "Analyse de risques à destination des non-spécialistes" [1] :
https://risques-tracage.fr/
https://risques-tracage.fr/docs/risques-tracage.pdf
nous pensons que notre solution permet de renforcer la sécurité du système. Nous publions donc cette proposition afin de tenter de faire avancer le schmilblick …

(1) Petite remarque à l'attention des auteurs du document d'analyse de risques (INRIA, CNRS, University of Waterloo, Sorbonne Université et EPFL) :

A la lecture des scénari que vous avez élaborés, il semble que vous soyez beaucoup plus impliqués, imaginatifs et efficaces que les stratèges de nos services publics en charge de la « veille et sécurité sanitaire » et de la « vigilance quand aux risques infectieux émergents ». Peut-être faudrait-il à l'avenir, que vos équipes soient consultées sur le long terme pour l'élaboration de scénari dans ces tâches d'anticipation et de préparation aux crises ;-)

Les spécificités de notre proposition sont les suivantes :

- Les codes uniques permettant d'identifier chaque rapprochement entre 2 appareils sont générés par l'application installée sur chaque appareil client.
- Seuls les codes uniques possiblement contaminants sont stockés sur la plateforme centrale. Soit au total tout au plus 10 millions de codes (pour la conservation durant 10 jours des codes de 100 000 personnes déclarées positives).
- Lors de l'interrogation quotidienne de la plateforme centrale par un appareil client, des codes fictifs d'offuscation sont ajoutés aux requêtes. Ces codes fictifs ne peuvent pas être distingués de codes liés à un rapprochement réel entre 2 appareils et rendent vain toute exploitation en dehors du protocole. Ce simple ajout de données fictives permettrait certainement de rassurer les utilisateurs quand à la confidentialité des informations communiquées et favoriserait l'adoption de cette application.
- Dans notre proposition, seul le transfert des trames entre clients et serveurs est chiffré (HTTPS). Ceci permet d'alléger la charge des serveurs qui prennent en charge le flux principal. Ces serveurs ont juste à répondre par vrai/faux après comparaison des codes soumis (par paquets) par le client avec ceux présents dans la liste des codes possiblement contaminants. Cet hébergement serveur sans base de données (les données étant limitées à une liste de codes) serait certainement beaucoup moins lourd et onéreux que celui choisi pour l'application StopCovid.
- Le flux des codes possiblement contaminants remontés par un appareil suite à une déclaration positive au Covid-19 seront acheminés par l’intermédiaire d'un VPN de confiance afin de masquer l'IP client à la plateforme centrale. Afin de rendre vain tout espionnage du trafic visant à identifier les appareils clients déclarant une contamination, chaque appareil client se connectera de façon aléatoire tous les 2 ou 3 jours à ce VPN en transmettant des paquets de codes reconnus comme nuls par la plateforme centrale.
- Une contamination déclarée à la plateforme doit être validée par un intervenant autorisé (le médecin, covidom.fr, …).

Si vous êtes intéressé par ce sujet, merci de ne pas vous arrêter à cette présentation succincte et de lire la description technique qui suit. Si vous êtes un spécialiste de la sécurité de ce type d'application, merci de nous faire part de vos remarques ou de nous dire éventuellement pourquoi il est totalement délirant de vouloir se passer de crypto pour ce genre d'application.

Lexique :

- *Appareil client :* le téléphone, smartphone, tablette ou autre appareil mobile (ou fixe [2]) sur lequel l'application cliente liée au Covid-19 a été installée.
- *GUID :* l'identifiant unique est (dans le cas de cette application) un code permettant d'identifier un évènement lié au rapprochement entre 2 appareils client.
- *GUID possiblement contaminants* : est un GUID émis par l'appareil d'une personne qui a été à postériori diagnostiquée/testée positive au COVID-19.
- *Plateforme centrale :* l'ensemble des ordinateurs qui hébergent les serveurs centralisant la liste des GUID possiblement contaminants ainsi que les listes de GUID en attente de confirmation par une autorité.
- *Bluetooth :* c'est une norme de communication à courte distance permettant l'échange bidirectionnel de données entre les appareils électroniques [3].
- *Code de session :* code numérique à 6 chiffres permettant à la plateforme centrale de regrouper un ensemble de GUID en attente de validation par un intervenant autorisé.
- *Offuscation* : https://www.pourlascience.fr/sr/logique-calcul/loffuscation-ou-lart-de-brouiller-lecoute-18265.php

La description en 4 points.

1) À chaque rencontre entre 2 appareils (détectée par Bluetooth) :

- Chacun des 2 appareils génère un identifiant unique [4] pour cet évènement, l'échange avec l'autre appareil et mémorise l'identifiant émis et l'identifiant reçu [5].

2) Lorsque Alice présente des symptômes du Covid-19 :

- Alice signale sa possible contamination à la plateforme centrale. L'application cliente installée sur son appareil y associe la liste des GUID qu'elle a émis au cours des 15 derniers jours [6].
- La plateforme mémorise ces GUID et l'associe à un code de session.
- Alice reçoit en retour le code de session unique généré par la plateforme.

3) Quelques jours (ou quelques heures) plus tard, lorsque Alice est diagnostiquée positive (par un test ou par un médecin) **:**

- Alice signale le diagnostic positif à la plateforme centrale. L'application cliente installée sur son appareil y associe la liste des GUID qu'Alice a émis depuis sa première déclaration.
- Alice reçoit en retour un code de session unique généré par la plateforme.
- Un intervenant autorisé (le médecin, covidom.fr, ...) valide sur la plateforme centrale (avec identité et certificat à l'appui) les 2 codes de session reçu par Alice.
- La plateforme centrale ajoute à sa liste les GUID possiblement contaminants associés à ces 2 codes de session.
- Alice interroge l'état des codes de session et reçoit un compte rendu de la prise en compte de sa contagion.

4) Chaque appareil client interroge la plateforme centrale régulièrement au cours de la journée [7] pour savoir si l'un des GUID reçu a été validé comme « possiblement contaminants » sur la plateforme.

- Les GUID reçus lors de rencontres sont envoyés à la plateforme par paquets de 100 (choisis au hasard dans la liste mémorisée sur l'appareil [8]).
- La réponse POSITIF/NEGATIF est attendue pour chaque paquet. Une réponse POSITIVE sera signalée immédiatement à l'utilisateur de l'appareil.
- Afin, entre autres, de rendre vain l'exploitation par la plateforme centrale de ces informations de façon détournée, des paquets d'offuscation sont également envoyés. Ces paquets contiennent des GUID fictifs générés aléatoirement [9]. Une éventuelle (et très peu probable) réponse positive pour l'un de ces paquets d'offuscation sera bien entendu ignorée.

(1)

(2) L'appareil fixe (balise Beacon programmable ou smartphone) pourrait être envisagé comme outil de pré-alerte dans un lieu sensible devant être préservé de la contamination (par exemple un service de maternité) ou dans des lieux où se croisent de nombreuses personnes (par exemple bars et restaurants).

(3) Cette norme Bluetooth de communication permettant l'échange bidirectionnel de données est implémentée aujourd'hui sur la plupart des appareils électroniques communicants. Toutefois, dans un soucis de sécurité, cette fonctionnalité n'est pas systématiquement activée sur les systèmes d'exploitation. D'autre part, d'après notre faible connaissance du sujet, un appairage est nécessaire avant tout échange de données entre 2 appareils !

https://fr.wikipedia.org/wiki/Bluetooth#Mise_en_%C5%93uvre

Cet appairage serait-il possible sans intervention de l'utilisateur ?

(4) Identifiant global unique (GUID) généré par l'application installée sur l'appareil client :

Chaque GUID est généré automatiquement à la demande par concaténation des infos suivantes :

- l'horodatage à la microseconde

- une info liée au contexte de l'appareil (par exemple nombre octets mémoire disponibles sur l'appareil modulo 10^8)

ou une info issue par exemple d'un texte d'une page web choisie au moment de l'installation de l'application sur l'appareil

Rmq : La génération de 2 GUID identiques sur 2 appareils distincts est très peu probable.

Si toutefois cela se produisait ça poserait problème si et seulement si un seul des GUID est remonté à la plateforme. Ce cas génèrerait un faux positif (parmi d'autres faux positifs issus de mauvais diagnostic ou mauvais tests ...)

(5) La mémorisation d'un GUID sur l'appareil, pourrait être enrichi des attributs suivants :

- la qualité du signal qui peut être le reflet de la distance entre les antennes Bluetooth (avec réserve quand à la fiabilité de cette information).

- la durée de l'approche entre les 2 Bluetooth.

- un tag « douteux » pour un GUID reçu qui ne soit pas compatible avec l'instant de l'approche Bluetooth. Un GUID douteux peut provenir d'un appareil client qui n'est pas à l'heure ou d'une application frauduleuse qui tente de recycler un GUID déclaré possiblement contaminant, dans le but de porter préjudice à d'autres utilisateurs.

- on peut également y ajouter un contexte "transport automobile, autocar, métro, train, ..." si l'accès aux capteurs de l'appareil a été autorisé.

Ces attributs complémentaires permettraient d'effectuer une interrogation prioritaire pour les GUID reçus proches et de longue durée.

Dans le but d'alléger la charge des serveurs de la plateforme centrale, les nombreux GUID reçus dans une proximité de courte durée pourraient être interrogés de façon étalée sur 3 ou 4 jours.

(6) La remontée à la plateforme des GUID émis par Alice est effectuée par paquets de 100 GUID (pour que les trames ressemblent aux trames d'interrogation périodique). Une taille aléatoire fixée entre deux limites permettrait d'ajouter de la discrétion vis à vis des opérateurs réseaux en charge du transport de la trame dans les échanges avec la plateforme.

A chaque paquet, la plateforme centrale accuse réception par l'envoi d'un identifiant paquet (que l'appareil d'Alice mémorise).

Après avoir remonté le dernier paquet et reçu le dernier accusé réception, l'appareil d'Alice remonte un paquet final contenant les identifiants des accusés réception mémorisés. La plateforme centrale regroupe alors tous les GUID émis par Alice sous un code de session.

Alice reçoit en retour ce code de session unique généré par la plateforme centrale pour la déclaration de sa contamination.

Il est évident qu'après cette déclaration, le comportement d'Alice doit être civique. Elle doit se confiner afin de ne pas risquer de contaminer d'autres personnes. Pour parfaire le système, on peut toutefois envisager la possibilité d'une nouvelle déclaration à l'initiative d'Alice, peu de temps avant la confirmation par le médecin.

L'intervenant autorisé aurait donc dans ce cas 2 codes de session à valider. Le premier pour les GUID émis par Alice 15 jours avant sa première déclaration et le 2ième code de session pour les GUID émis par Alice entre les 2 déclarations.

(7) Au cours de la journée, chaque appareil réserve une période aléatoire de quelques heures sans aucune communication avec la plateforme. Ceci permet lors des déplacements que les opérateurs réseaux en charge du transport des trames ne puissent pas identifier à coup sûr, les appareils équipés ou non de l'application.

(8) Un envoi ordonné des GUID reçus n'est pas souhaitable car permettrait de connaitre la date et l'heure de contact possiblement contaminant. Ce qui en cas de hack coté client, permettrait éventuellement d'identifier le contact contaminé.

(9) Les GUID fictifs sont générés aléatoirement chaque jour et contiennent bien entendu un horodatage compatible avec la journée concernée. L'application mémorise ces GUID fictifs afin de les réutiliser les jours suivants et de simuler ainsi un comportement identique aux GUID réels.

Le nombre de paquets d'offuscation sera plus important pour un appareil client qui émet et reçois très peu de GUID. Ceci permet de masquer aux opérateurs réseaux le profil du client (confiné, ayant de nombreux déplacements dans des environnements denses en population, ...).

Qui sommes nous :

Bob : automaticien industriel et informaticien
Thierry : retraité 👍

N'hésitez pas à nous faire parvenir vos remarques : contact@bobiciel.com

C'est étrange cette paranoïa qui semble se développer chez les utilisateurs de smartphone ! Alors que depuis plus d'une décennie, nos vies ~~numériques~~ sont (avec l'accord de ces utilisateurs) suivies, tracées, profilées, calibrées, triées, sélectionnées, croisées, analysées, exploitées, … par entre-autres les Google, Amazon, Facebook, Apple, Microsoft, LinkedIn, Booking, ... sans que ça émeuve les foules ...

Version 1.4

Juillet 2020

Notes de versions

Nous remercions nos lecteurs pour leurs questions et leurs retours.

V 1.3

- Ajout de la définition « Code de session » au lexique.
- Ajout aux spécificités de notre proposition d'une remarque concernant le chiffrement.
- Nous ne pensons pas que le timestamp intégré au GUID soit une faille. Car de toute façon cet horodatage est indispensable coté client. Coté plateforme centrale, à partir du moment ou plusieurs milliers d'utilisateurs remontent des GUID (réels et fictifs), il nous semble impossible d'exploiter ces GUID autrement que prévu par le protocole.

  De plus, cet horodatage intégré aux GUID permet à l'application cliente de taguer « douteux » un GUID reçu qui ne soit pas compatible avec l'instant du rapprochement Bluetooth *voir (5)*.
- Amélioration de la génération des GUID fictifs *voir (9)* afin de les rendre indécelables coté serveur.
- Suite aux échanges avec Jean, les améliorations suivantes pourraient être apportées :
  - Lorsque l'application cliente reçoit une réponse positive pour un paquet de GUID, elle tague ces GUID et n'interroge plus la plateforme pour ces GUID.
  - Coté plateforme, un compteur est associé à chaque GUID possiblement contaminant afin de s'assurer que ce GUID ne soit interrogé qu'une seule fois. Ce comptage permettra de diagnostiquer coté serveur, un requêtage anormal sur certains GUID possiblement contaminants.

V 1.4

- Réécriture du § « Les spécificités de notre proposition »