# Blogpost: Evaluating COVID-19 contact tracing apps? Here are 8 privacy questions we think you should ask.

*Yves-Alexandre de Montjoye, Florimond Houssiau, Andrea Gadotti, Florent Guepin*

*Transcript of the blogpost at https://cpg.doc.ic.ac.uk/blog/ evaluating-contact-tracing-apps-here-are-8-privacy-questions-we-think-you-should-ask/*

**Abstract** – *While governments are ramping up their efforts to slow down the spread of COVID-19, contact tracing apps are being developed to record interactions and warn users if one of their contacts is later diagnosed positive. These apps could help avoid long-term confinement, but also record fine-grained location or close-proximity data. In this blog post, we propose 8 questions one should ask to understand how protective of privacy an app is.*

---

*April 2, 2020*

As strong measures are being put in place to slow down the spread of COVID-19, many are looking at how technology and data could help. With many countries using mobile phone location data [1,2,3,4,5] to analyze the effectiveness of social distancing measures and help predict the potential geographic spread of the disease, the focus has now shifted to whether mobile phones could also help warn users if they have been exposed to an infected person.

Contact tracing apps are being developed in the UK [6], US [7], Germany [8], with one already deployed in Singapore [9]. These apps have increasingly come under the spotlight as a potential long-term way to monitor the virus. Epidemiologists say that it could prove vital [10, 11] to avoid long-term extreme confinement measures. The data handled by these apps, from location data [12] to fine-grained close proximity information [13] and whether a person might be infected and should self-quarantine is however very sensitive.

In our Mar 21 blog post [14], we emphasized how we do not have to pause privacy laws and regulations [15] and how privacy engineering can help measure and limit the spread of the virus without resorting to mass surveillance.

When it comes to contact tracing, this however requires to go beyond simple re-assurances that the phone numbers aren't recorded, that everything is encrypted, that pseudonyms are changing, or that the app is based on consent. Indeed, a large range of techniques exist to circumvent those protections. For instance, scores [16] have been developed to re-identify individuals in location [17] or graph datasets [18, 19], session fingerprinting [20] could be used to link a pseudonymous app user to an authenticated web visitor, and node-based intrusions would allow to track users [21].

Apps are being developed around the world and are likely to be available within weeks. If they are proven useful, governments, health authorities, and users will have to evaluate the different approaches and decide whether to adopt them.

Privacy is a crucial component in the equation. In this post, we propose 8 questions one should ask to assess privacy in contact tracing apps.

## Contact tracing – Setup

We here focus on contact tracing apps installed by users and empowering them: apps informing users that they have been in close proximity with an infected individual in the past and providing them with recommendations on what to do.

The privacy setting for such an app typically involves the following entities:

- **Users** who install the app on their phone;
- **Authority** (e.g. the government or a healthcare provider) that runs a central server coordinating the contact tracing;
- **External entities**: malicious apps, users, a foreign agency, or a company who is trying to take advantage of the situation to collect data or tamper with the contact tracing.

We use *user* to refer to anyone using the app, *infected* for users who have been tested and were found to be positive, and *at risk* for users who have been in close contact with someone who was later found to be infected.

A digital contact tracing app would typically work like this: Bluetooth signals or location information are recorded by phones; when a user is diagnosed positive (*infected*), they upload their data (under some form) to the authority, which then arranges for other users to learn that they are *at risk* because they interacted with an infected person.

To illustrate the vulnerabilities our questions are meant to surface, we use three "toy" protocols. Note that they are only meant to illustrate the questions and are not complete, deployable solutions for contact tracing, nor even good protocols.

*Toy protocol 1 (using location):*

- Each app only records its own location.
- When a user reports as infected, they send their trajectory (location and time) to the authority.
- The authority shares the pseudonymous trajectories of all infected users with every user. Users can then check if they were in close contact with an infected individual.

*Toy protocol 2 (using Bluetooth):*

- Each app broadcasts a unique identifier assigned by the authority through Bluetooth.
- When two phones are near to one another, they exchange these identifiers.

- When a user reports as infected, they send all the identifiers they encountered to the authority.
- The authority contacts all the users whose identifier was encountered by an infected user.

*Toy protocol 3 (using Bluetooth):*

- Each app broadcasts a unique identifier using Bluetooth, assigned by the authority. This unique identifier is reset every hour.
- When two phones are near to one another, they exchange these identifiers.
- When a user reports as infected, they send all the identifiers that they have used (one per hour) to the authority.
- The authority shares the identifiers of all infected users with every user. Users can then check if they encountered one of these identifiers recently.

Using this vocabulary and definitions, we propose 8 privacy questions that we would like app developers to answer. We hope these questions will help start a high-level discussion to systematically evaluate potential vulnerabilities and real risks in existing and future contact tracing apps.

## The questions

### 1. How do you limit the personal data gathered by the authority?

*Large-scale collection of personal data can quickly lead to mass surveillance.*
In protocol 1, the authority learns the whole trajectory of infected users. In protocol 2, the authority learns the entire pseudonymous social graph of infected users, along with timestamps, which has been shown to be easily re-identifiable [18]. Both collect large amounts of personal data. Protocol 3 does better in that regard, with the authority only observing the pseudonyms of infected users (with changing identifiers).

### 2. How do you protect the anonymity of every user?

*Users' identities should be protected. Special measures should be put in place to limit the risk that users can be re-identified by the authority, other users, or external parties.*
Protocol 1 doesn't technically give the authority the identity of users. However, research [17] shows that location traces are highly unique, and could probably be easily linked back to a person. Protocol 2 is worse, as the users are given a unique identifier by the authority, which can link these identifiers to the phone. Protocol 3 is much better – as long as connections with the server are anonymous (e.g. using Tor [22] or mixes [23]), the user's identity could be kept secret.

### 3. Does your system reveal to the authority the identity of users who are at risk?

*The goal of contact tracing is to let people know they have been in contact with someone who was infected. The authority should not know who these people are.*
No for protocols 1 and 3, which never require data from non-infected users. Yes for Protocol 2, in which the authority explicitly contacts at risk users, and could use this information to, e.g., force them into quarantine.

**4. Could your system be used by users to learn who is infected or at risk, even in their social circle?**

*Having been in contact and infecting someone may become a matter of life and death. Digital contact tracing should warn people at risk without revealing who might have infected them.*
Protocol 1 exposes the full data of infected users publicly: every user can then check if a particular person they know is in the dataset. In protocol 3, a user at risk learns the hour at which they met an infected user, and can probably find out who infected them. Protocol 2, on the other hand, prevents users from learning anything about one another.

**5. Does your system allow users to learn any personal information about other users?**

*Apps should not need to leak information on a user's locations or social networks to other users.*
All three protocols protect data of non-infected users, but only protocol 2 prevents users from learning anything about infected users. Protocol 1 exposes their entire trajectory. Protocol 3 leaks small amounts of information: identifiers encountered by infected individuals. It is possible for a user to recognize identifiers they have encountered and learn that the user to whom the identifier belongs is at risk.

**6. Could external parties exploit your system to track users or infer whether they are infected?**

*The system should take into account the risk of external adversaries, including well-resourced ones.*
Both protocols 2 and 3 force phones to broadcast an identifier. An entity could install Bluetooth trackers to cover a city, or install malicious code on phones, and record the identifiers that they observe in specific locations. In our research [21], we showed that trackers installed on the phones 1% of London's population would allow an attacker to track the real-time location of over half of the city. Protocol 3 makes this attack much more difficult, as the identifiers change every hour.

**7. Do you put in place additional measures to protect the personal data of infected and at risk users?**

*The system design may require revealing more personal information about users who are infected or exposed. But these are often the people who are more vulnerable and at risk.*
Protocol 1, and to some extent protocols 2 and 3, require infected users to share more data. Users at risk are however safe in protocol 1 and 2, but not in protocol 3, where some of the identifiers that they have used are published.

**8. How can we verify that the system does what it says?**

*Large-scale contact tracing is too sensitive to rely on blind trust. Transparency is essential to prove that the app functions as advertised.*
Transparency of the full system is absolutely fundamental to guarantee privacy. This requires open protocol specifications, but also public source code, reproducible builds [24], and verifiability of what is being broadcasted by apps.

---

In the next few weeks, most of us are likely to install a contact tracing app to help slow down the spread of coronavirus. The questions we propose here represent a starting point for an informed conversation on the privacy risks of apps we are being offered.

Importantly though, they cannot replace a full independent privacy audit. In-depth formal analysis of the protocol is necessary before deployment and should be published. Protecting privacy should rely on mathematical proofs of correctness, with mitigation strategies considered only when necessary. Our questions focus on privacy aspects, but ensuring security is similarly crucial. This means, for example, supervising the integrity and authenticity of the crowdsourced data, evaluating how mobile malware could affect the app's behavior, or assessing the resilience of the authority's servers against intrusions.

Building a contact tracing app that allows all of us to participate in the fight against COVID19 is possible, but it will require us to go beyond shallow reassurances that privacy is protected.

---

## Bibliography

[1] A. Cloot (2020), 'Coronavirus: le cabinet De Block dit «oui» à l'utilisation des données télécoms', *Le Soir*, March 12

[2] M. Sweney and A. Hern (2020), 'Phone location data could be used to help UK coronavirus effort', *The Guardian*, March 19

[3] B. Fung (2020), 'Trump administration wants to use Americans' location data to track the coronavirus', *CNN Business*, March 18

[4] M. Scott, L. Cerulus and L. Kayali (2020), 'Commission tells carriers to hand over mobile data in coronavirus fight', *Politico*, March 23

[5] J. A. Gross and T. Staff (2020), 'Israel starts surveilling virus carriers, sends 400 who were nearby to isolation', *The Times of Israel*, March 18

[6] R. Manthorpe (2020), 'Coronavirus: Govt set to release 'contact tracking' app which detects nearby virus carriers', *Sky News*, March 31

[7] Raskar, R., Schunemann, I., Barbar, R., Vilcans, K., Gray, J., Vepakomma, P., Kapa, S., Nuzzo, A., Gupta, R., Berke, A. and Greenwood, D., 2020. Apps gone rogue: Maintaining personal privacy in an epidemic. arXiv preprint arXiv:2003.08567.

[8] D. Busvine (2020),Germany aims to launch Singapore-style coronavirus app in weeks, *Reuters*, March 30

[9] The TraceTogether app: https://www.tracetogether.gov.sg/

[10] Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Bonsall, D.G. and Fraser, C., 2020. Quantifying dynamics of SARS-CoV-2 transmission suggests that epidemic control and avoidance is feasible through instantaneous digital contact tracing. medRxiv.

[11] University of Oxford (2020), 'Infectious disease experts provide evidence for a coronavirus mobile app for instant contact tracing', *University of Oxford news*, March 17

[12] J. Valentino-DeVries, N. Singer, M.H. Keller and A. Krolik (2018), 'Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret', *The New York Times*, December 10

[13] Privacy International (2020), 'Bluetooth tracking and COVID-19: A tech primer', *Privacy International*, March 31

[14] Y.-A. de Montjoye, F. Houssiau (2020), 'Blogpost: Can we fight COVID-19 without resorting to mass surveillance?', *CPG Blog*, March 21

[15] S. Mullainathan and R. Thaler (2020), 'To Fight the Coronavirus, Cut the Red Tape', *The New York Times*, March 24

[16] Riederer, C., Kim, Y., Chaintreau, A., Korula, N. and Lattanzi, S., 2016, April. Linking users across domains with location data: Theory and validation. In Proceedings of the 25th International Conference on World Wide Web (pp. 707-719).

[17] De Montjoye YA, Hidalgo CA, Verleysen M, Blondel VD. Unique in the crowd: The privacy bounds of human mobility. Scientific reports. 2013 Mar 25;3:1376.

[18] Narayanan, A. and Shmatikov, V., 2009, May. De-anonymizing social networks. In 2009 30th IEEE symposium on security and privacy (pp. 173-187). IEEE.

[19] Sharad, K. and Danezis, G., 2014, November. An automated social graph de-anonymization technique. In Proceedings of the 13th Workshop on Privacy in the Electronic Society (pp. 47-58).

[20] Alepis, E. and Patsakis, C., 2018. Session Fingerprinting in Android via Web-to-App Intercommunication. Security and Communication Networks, 2018.

[21] Radaelli, L., Sapiezynski, P., Houssiau, F., Shmueli, E. and de Montjoye, Y.A., 2018. Quantifying surveillance in the networked age: Node-based intrusions and group privacy. arXiv preprint arXiv:1803.09007.

[22] Dingledine, R., Mathewson, N. and Syverson, P., 2004. Tor: The second-generation onion router. Naval Research Lab Washington DC.

[23] Chaum, D.L., 1981. Untraceable electronic mail, return addresses, and digital pseudonyms. Communications of the ACM, 24(2), pp.84-90.

[24] moxie0 (2016), 'Reproducible Signal builds for Android', *Signal blog*, Mar 31