On the Privacy of TraceTogether, the Singaporean COVID-19 Contact Tracing Mobile App and Recommendations for Australia

Authors (in Alphabetic Order):

Hassan Asghar (Macquarie University), Farhad Farokhi (University of Melbourne),
Dali Kaafar (Macquarie University), Ben Rubinstein (University of Melbourne) [1]

The Australian government is exploring the use of contact tracing mobile apps as a tool for public health officials and communities to fight the spread of the COVID-19 pandemic. "Digital methods can be also used to assist in identifying contacts and to be able to shut those (tracing) issues down as was practiced in Singapore. And so these delivery methods are also being looked at by the Commonwealth, and we're making a lot of progress there." Scott Morrisson, Prime Minister of Australia March 24, 2020.
Earlier this month, as one of the countries that has so far managed to keep the pandemic under control, Singapore announced the release of mobile app "TraceTogether." The app tracks the interactions between those diagnosed with coronavirus and the wider community. Australia is currently fast-tracking the review process for TraceTogether to be adopted and deployed across the country.

There are however privacy concerns and implications if TraceTogether or similar apps are to be deployed in Australia. While many of the legal considerations could be relaxed at the discretion of enforcement authorities during times of crisis (e.g., public health emergencies), privacy issues could seriously slow down the adoption of these mobile apps. An additional consideration is that such apps could be used as a tool for mass surveillance beyond its original purpose of COVID-19 contact tracing.

Time is critical in the fight against the spread of COVID-19, and so building a contact tracing app from scratch, with a completely new design, is not an option. Fortunately Singapore has announced its intention to open-source the TraceTogether app. We take the pragmatic view of assessing the privacy of TraceTogether while providing recommendations on how its privacy

[1] The authors thank Vanessa Teague for fruitful discussions.

can be enhanced without drastically changing its design and hence enabling timely implementation. These recommendations can be implemented as enhancements to the current TraceTogether app's code.

A Review of TraceTogether:

TraceTogether relies on Bluetooth to exchange information between TraceTogether users including Bluetooth signal strength (a proxy for distance between users), time, and (temporary) user IDs. This information is locally logged (on-device) by the app in an encrypted form. When a user installs the TraceTogether app, they submit their mobile number to the centralised authority (Ministry of Health server in the case of Singapore). This mobile number together with a newly generated user ID is stored on the server. The server, using its private key, generates temporary IDs, which are periodically refreshed, and transmits them to the corresponding user. These temporary IDs are exchanged between TraceTogether users when in close proximity.

When a TraceTogether user is diagnosed with COVID-19, they are asked for consent to upload the app's encrypted data logs to the server. The data logs are then decrypted by the server (having the private key). Through these data logs the temporary IDs in the logs can then be used to contact other TraceTogether users who were in contact with the infected user. Since these temporary IDs are generated by the server, the server can look them up against its user ID-mobile number database to determine the identity of these users.

Other countries have also developed similar apps for tracing the spread of COVID-19. For instance, Israel has released Hamagen (The Shield) that retains time and location information on mobile devices and cross-references this information with the Israel Ministry of Health's updated epidemiological data (i.e., time and location of infected people). MIT Media Lab has also released an open-source app, called Safe Paths, that works similarly based on location history.

Summary of Privacy Features:

We can think about privacy threats from three entities: other TraceTogether users, the server or central authority (e.g., the government), and snoopers (e.g., malicious users, scammers, hackers).

The TraceTogether app provides privacy from other users: The app generates temporary user IDs and refreshes them frequently. These IDs are generated by the server based on the phone numbers of people and their permanent ID so that the central authority can determine the identities of users if needed. The temporary or varying nature of the IDs implies that users can not be traced over a long period by other users. However, the server can potentially do this.

The app keeps the users safe from snoopers: The data logs on the user's phone are encrypted. So, if anyone hacks into the user's phone, the data is not intelligible. The server has the key to decrypt the data logs, which are only sent to the server for determining close contacts between people.

The app does not provide sufficient privacy from the Central Authority: The server, after retrieving the data log of the users, can decrypt and read it. The server can also link the temporary IDs to the real identity of users. However, some privacy features are present: The server does not ask for the data log of the users who are not infected or have not been in close proximity of an infected user. The data log only contains relative distance (via Bluetooth signal strength), and not the exact location where the users come in close contact. The data on the phones (not the data transmitted to the servers) is deleted after 21 days.

The server can know the private data of a user even if they are not infected: Once a user has tested positive for COVID-19, they need to provide consent to the server to retrieve and decrypt their data log. This enables the server to obtain the identities of the TraceTogether users that have been in contact with the infected user. These potentially uninfected users are no longer in control of their privacy.

The app can potentially be (mis-)used for surveillance: Even though the data logs are only sent to the Central Authority following user's consent, there is no check to ensure that the request from Central Authority is genuine or not, i.e., whether that user was in proximity of an infected user. Thus, a curious Central Authority might be able to obtain and decrypt data logs from a large number of users yielding to potential mass-surveillance threat. Furthermore, even though the local on device data logs are deleted after 21 days, there is no guarantee that the data logs decrypted at the Authority server would also be deleted.

Privacy Recommendations:

The app can be tweaked to provide more privacy from the Central Authority: The temporary user IDs can be generated locally by the app, instead of the server. This way, no one except the users know their identity, and they have to provide an informed consent to the server by sharing the list of their temporary IDs for their private data to be used. When a user is diagnosed, the server can find the temporary IDs that have been in contact with the infected user and broadcast them. When they receive a message that contains their temporary ID, users can respond by identifying themselves. This functionality can be implemented either as a specific consent request to individuals or as an automatic response to identification requests if individuals opt in for self-identification.

Future versions of the app need to be more decentralized: The server can push the temporary ID of diagnosed users to the apps and other users can locally determine if they have been in contact with them. If the IDs are locally and randomly generated, they are not linkable to

true identities. This way the server does not know the identities of the users who were in close proximity with the infected user. This provides higher privacy against the server. This is not an easy fix and requires a fundamental change in the app's design which might not be possible to be implemented rapidly.

Future releases of anonymised data logs must be restricted: An important aspect of data gathered by the server is future use by epidemiologists and policy makers. Although the information seems innocuous, it can be very sensitive and reveal a lot about the users. So it should not be shared publicly even if anonymized. This is because a large percentage of the people might share their data. Even the contact graph, without locations, timestamps, phone numbers or explicit identities, can be linked to other data sources enabling user reidentification.