Privacy at the time of pandemic

Stan Matwin, Ph.D., CRC
Institute for Big Data Analytics
Dalhousie University,
Halifax, NS, Canada
stan@cs.dal.ca

"It may be necessary temporarily to accept a lesser evil, but one must never label a necessary evil as good."

– **Margaret Mead**

"The ultimate test of man's conscience may be his willingness to sacrifice something today for future generations whose words of thanks will not be heard."
– **Gaylord Nelson**

This is a draft working paper spawned by very recent developments we are all experiencing, and by discussions with several colleagues.

1. It seems that epidemiology and infectious disease experts agree that, in the absence of a vaccine or treatment, the only course of action is first confinement, such as we are experiencing now, eventually gradually relaxed to prevent a total economic collapse; which is then followed by the track-and-trace process. Track and trace means massive testing (ideally 100% of the population), followed by tracking and isolating quickly all the contacts of the people who tested positive. Some eminent economists argue that the economic effects of the broad, non-specific isolation policy will be so nefarious for the economy that we should consider replacing it with testing and only isolating the positive cases [10]. Their simulations indicate that, in terms of how many people get infected, the effect of such selective isolation, in terms of the spread of the infection, will be comparable to the current quarantine policies followed by most countries. At the same time, many fewer people will be isolated, alleviating the damages to the economy. Testing, however, must be followed by tracking and tracing in order to identify those infected by a tested positive person. If testing does not cover 100% of the population, and it never will, tracking and tracing is necessary to locate potentially untested contacts of the known positives[1].
2. Tracking-and-tracing is an old procedure, used routinely in past epidemics. The idea is to identify all the contacts: people with whom the infected, positive person interacted over a defined period of time, then in turn to identify their contacts, etc. The period of time and the precise definition of "interaction" (e.g. the distance between the positive person and the contact) are parameters of the tracking procedure. The contacts are found by intersecting spatially and temporally the human mobility information. In the past tracking was done manually by teams of

[1] The cited proposal suggests testing 7% of the population per day, so that every person would be tested every 15 days. But that would mean 20M people per day tested in the US, or 2M in Canada. These numbers far exceed the capacities of the existing infrastructure in terms of the number of test available and the personnel interpreting the results. Availability of an automated test with immediate results would resolve this.

investigators interviewing chains of people, all starting with the positive person. However, in a pandemic with an R0 factor as large as in COVID-19, the manual methods are unfeasible[2].

3. An obvious modern tool providing information about human mobility is the cellphone. Given the scale of the pandemic, this is also the only way. There are several goals the public health system may want to achieve using data from mobile phones. This means that there are several variants of tracing and they differ in what data is collected and how. Such variants come with different privacy exposures. An easy and obvious one is to monitor people subject to a strict 14-day self-isolation. Should such a person leave the location they are confined to, the tower receiving their mobile phone signal will almost certainly change (especially in an urban setting, where distance to the tower is often a few hundred meters), which can be easily and automatically detected by the existing mobile service providers, without any additional infrastructure. Another goal is pursued by an app recently proposed by Yoshua Bengio [1] that would inform an individual about a risk they run by contacting another person, also a holder of the app. The proposal is a fully distributed architecture without the need of a centralized, trusted database. It seems that the goal of this proposal is different from the one pursued by the track-and trace approach: the idea is "…not to blame or identify, but rather to provide citizens with the information they need to minimize the risk of being contaminated" [1]. The goal of track-and-trace, however, is to identify all contacts of an infected person in the period (e.g. 14 days) preceding the time when the infection is determined. There are several proposals, but the one implemented in Singapore (a country with a 160% saturation of cellphones in the population – i.e. an average Singaporean owns 1.6 cellphones) is particularly well thought out and elegant in its design [2]. A similar design is mentioned in connection with the joint project by Google and Apple. Ideally, people's participation in such a system would be voluntary, but there is no evidence outside South-East Asia that this approach will ensure sufficiently complete participation of the population. Notwithstanding how well it does its job, it poses some important questions about personal data and their use by the government.
4. Yet another public health goal, implying another variant of track and trace, is raised by those that emphasize the need to detect clusters of infections: events co-located in time and space in which lots of future positives participated. It is believed that such clusters, also known as "superspreaders", are responsible for large proportion of infections, see eg. [13]. To detect such clusters, person's location and time data would have to be open to the authorities, potentially further exposing their privacy. Again, as there is significant potential for highly effective tracking and tracing, we should consider opening that data. In some cases as this could be done by people volunteering their data, but collecting it universally would result in a much more robust tracing. This would also need to be constrained, e.g. by the time length when such surveillance is data is being collected, and the switching-off should be verifiable externally in a trusted manner (see p. 8 below).
5. The Singapore system, called TraceTogether (see [3] for a brief description), is based on the idea that a contact takes place between two people if their cellphones are close enough to communicate by Bluetooth. Nearby phones exchange digital tokens. Tokens are associated with a person in an encoded manner, and they remain in the phone which has received them. Tokens

[2] Media reports e.g.([11]) indicate that Canada is considering putting in place a "manual" track and trace system. Recent experience from other countries indicates that this may not work [12].

of an individual are changing over time, and this is an important design feature. When a positive person is identified, that person is legally obliged to release all the tokens their cellphone has received over a time period (eg 14 days) to the health authorities (in this case, the Singapore Ministry of Health, or SMH). SMH can decode the tokens and link them to phone numbers and specific persons. In Singapore, users install TraceTogether on a voluntary basis and the same voluntary approach is part of the Google&Apple design [8]. However, in the case of the latter, after the initial period the app will be built into the OS of the phone and therefore will become compulsory.

6. Obviously, privacy is an issue with TrackTogether. On the one hand, in research, the generally accepted privacy framework today is differential privacy (DP) [4]. DP postulates that a database is differentially private if any query asked of that database is oblivious to a presence of a particular individual in that database. This is contradictory to the goal of tracking-and-tracing which is to identify a specific individual. On the other hand, in practical approaches to data privacy, especially in medical settings, the number one approach is *k*-anonymity. It generalizes the data so that any query will return an answer consisting of at least *k* records. This, again, does not work for tracking-and-tracing.
7. Paper [3] introduces a very simple taxonomy of privacy, appropriate for track-and-trace systems. It distinguishes three levels of privacy: privacy against a snooper, privacy from contacts, and privacy from the government. In TrackTogether, privacy between users is protected: the user cannot know to whom the tokens belong. Privacy from contacts is also protected; the well-known linkage attack [5] would not succeed because tokens change over time, sufficiently often to make this kind of attack impossible. There is no privacy from the authorities (SMH) in the current design of TrackTogether, but [3] shows how this could be engineered into the system. However, the proposed reinforcement of privacy - for the users to remain anonymous from the authorities - has a significant complexity and computational cost, due to the cryptographic tools used in that solution.
8. It is clear that the track-and-trace apps are antithetical to privacy. Their role, in the end, is to identify the infected people. Is this an acceptable goal? Should we accept this drastic privacy breach in the interest of the society? A recent IPSOS poll tells us that 65% of Canadians support the government's use of cellphones to track people in self-isolation. Personally, I answer "yes " to both questions. I think, however, that a responsible solution should augment the current designs of the track-and-trace systems. Additional technical features need to constrain the use of the collected data, while still making it entirely adequate for the goal of tracking-and-tracing. The obvious governance constraint is that such a system should be disbanded, i.e. uninstalled and deactivated at the software level, once the pandemic ends, e.g. when the whole population will have been vaccinated[3]. This requirement should be verifiable by system and code audit done by an independent body of a kind similar to the Electronic Frontier Foundation. Following that, a technical requirement having to do with the longevity of the data should be part of the design and implementation. The digital tokens should have a lifespan of 14 days, and should self-erase afterwards. This should be true of both tokens on the phones and tokens uploaded to

[3] It is likely that there may not be a "discrete", well defined end of the pandemic. It is therefore important that a threshold on the number of cases or another measurable, verifiable goal is stated upfront to end surveillance.

SMH. There are cryptographic techniques for addressing this requirement [6], but they are too complex and computationally too heavy to be useful inside a track-and-trace system, especially on mobile devices. This seems to be an important area for future research, perhaps involving integrated hardware/software solutions.

9. The goal of these remarks is to initiate a discussion between Computer Scientists, data scientists and computer ethicists interested in the privacy aspects of track-and-trace systems using human mobility data. Should we be revising our beliefs, at least temporarily? How can we meet the public health goals with methods that are simple, efficient, easy to implement, have minimal computational overhead and do what is needed? Should privacy be allowed to temporarily play second fiddle, with an understanding that there is a well-defined end to this relaxation of rules? What legislative steps are need in this new situation? HIPAA rules have been very significantly relaxed by the US Department of Health and Human Services [7]– what about other regulatory environments?

[1] Bengio, Y., Peer-to-peer AI-tracing of COVID-19 https://yoshuabengio.org/, March 23, 2020

[2] Singapore Government Blog: Help speed up contact tracing with TraceTogether https://www:gov:sg/article/help-speed-up-contact-tracing-with-tracetogether, March 2020

[3] Cho, H., Ippolioto, D., Yu,W., Contact Tracing Mobile Apps for COVID-19: Privacy Considerations and Related Trade-offs, https://arxiv.org/abs/2003.11511

[4] C. Dwork, A. Roth et al., The algorithmic foundations of differential privacy, Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, 2014

[5] L. Sweeney. k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002; 557-570.

[6] Geambasu, R. Kohno, T., Levy, A. Levy, H. Vanish: Increasing Data Privacy with Self-Destructing Data. USENIX Security Symposium 2009: 299-316

[7] Relaxing of HIPAA Laws During COVID-19 Pandemic, The M National Law review, March 18, 2020

[8] Contact Tracing – Bluetooth Specification, Google, Inc. Apr. 2020 https://covid19-static.cdn-apple.com/applications/covid19/current/static/contact-tracing/pdf/ContactTracing-BluetoothSpecification.pdf

[9] IPSOS, Apr. 2020, https://www.ipsos.com/en-ca/news-and-polls/Canadians-Supportive-Of-Wide-Ranging-Measures-To-Battle-COVID19-Including-Some-Surveillance

[10] Paul Romer's posts of March 23-25, 2020: https://paulromer.net/

[11] National Post, Apr. 12, 2020. https://nationalpost.com/news/politics/canada-looking-to-prepare-surge-force-use-cellphone-data-to-contain-covid-19).

[12] Le Monde, Apr. 8, 2020 https://www.lemonde.fr/societe/article/2020/04/08/coronavirus-la-france-sur-la-piste-de-son-patient-zero_6036012_3224.html

[13] How a Premier U.S. Drug Company Became a Virus 'Super Spreader'., N.Y. Times, 13/4/20. https://www.nytimes.com/2020/04/12/us/coronavirus-biogen-boston-superspreader.html