TIP
Tech Informed Policy

JUNE 4, 2020

COVID-19 RAPID TECHNOLOGICAL BRIEFING

Evaluation questions to assess a digital contact tracing/exposure notification application

Produced by

Taylor Owen, Policy Lead, Beaverbrook Chair in Ethics, Media and Communications, Director of the Centre for Media, Technology and Democracy, and Associate Professor in the Max Bell School of Public Policy, McGill University

Derek Ruths, Tech Lead, Director of the Network Dynamics Lab and Associate Professor of Computer Science, McGill University

Stephanie Cairns, Research Assistant

Sta Kuzviwanza, Research Assistant

Sara Parker, Research Assistant

Sonja Solomun, Research Director, Centre for Media, Technology and Democracy, McGill University

Kate Gilbert, Graphic Designer

Centre *for* MEDIA, TECHNOLOGY *and* DEMOCRACY

network dynamics @mcgill
measuring and predicting large-scale human behavior

ABOUT TIP

Tech Informed Policy (TIP) is an initiative spearheaded by two leading McGill researchers—Dr. Derek Ruths, Director of the Network Dynamics Lab and Associate Professor of Computer Science, and Dr. Taylor Owen, Beaverbrook Chair in Ethics, Media and Communications, Director of the Centre for Media, Technology and Democracy, and Associate Professor in the Max Bell School of Public Policy. TIP aims to demystify the technology underlying critical policy issues and to provide valuable, tech-based recommendations to Canadian policymakers.

For enquiries, please contact Derek Ruths.

Glossary of Terms

Application Programming Interface (API): An API provides a framework for developers to create their own programs. It is a collection of potential operations that programmers can develop to suit their needs.

Contact log: Contact logs are records which store the contact IDs of those with whom the user interacted, and other relevant information, like the time of contact.

Contact Tracing (CT): CT is the process of identifying exposed individuals who have come into contact with diagnosed individuals, as well as potentially notifying users of areas where there is a high risk of contracting COVID-19. CT requires the use of a central server, accessible by health authorities, to log the contacts of infected persons.

Exposure Notification (EN): An EN application logs the contacts of each user and stores this information on their individual devices, and not on a central server. If a user identifies themselves as testing positive for COVID-19, the users with whom they have come into contact are notified that they may have been exposed to the virus.

ID: The contact ID is a series of randomized characters broadcasted by a user's device throughout the day. Upon encountering another user, their IDs are exchanged and then stored in their respective contact logs. If a user later tests positive for COVID-19, their IDs from the previous 14 days are cross-referenced with those in other contact logs, so users who came into contact with the infected user can be notified.

Open source: Code (written in a programming language) that is "open source" is freely available to the public. An open source platform ensures that transparency, assuages public security concerns, and enables easier troubleshooting.

Token: The token generates each user's contact IDs. With EN, the user's device creates a new token every day, which then generates IDs that are broadcast throughout the day. With CT, the server creates a new token every day, and that token then generates IDs for every connected device.

EXECUTIVE SUMMARY:

The purpose of this briefing is to provide a framework and approach for evaluating proposals for Bluetooth exposure notification (EN) and contact tracing (CT) applications. The briefing provides guiding evaluative questions which are partitioned into two key themes: privacy and security, and adoptability and implementation. The proposed questions serve as a framework for assessing the effectiveness, feasibility, and potential limitations and harms of a given CT/EN proposal.

These guiding questions serve to ensure the use of CT/EN applications does not infringe upon Canadian citizens' fundamental rights and freedoms; that the technology under question is secure, transparent, verifiable, and used only to address public health purposes in slowing the rate of transmission of COVID-19 and advancing public health and safety during the COVID-19 pandemic.

In employing this framework to evaluate a specific CT/EN proposal, policymakers should demand thorough and transparent implementation details rather than vague assurances or technological promises. For example, a satisfactory answer to the question, "how can the app be customized to suit provincial and regional needs?" would explain, in detail, how the developer would work with provinces to modify the application, which aspects would be modified, how often, and by whom.

Contact tracing (CT) and/or exposure notification (EN) proposals and already implemented applications should be reconsidered under evolving conditions of their use, especially if an app:

- is deemed ineffective at slowing transmission rates or accurately informing users
- is demonstrably vulnerable to data misuse, security breaches, and/or malicious attacks
- is shown to cause undue harm and discrimination and/or infringe upon users' fundamental rights and freedoms

BACKGROUND

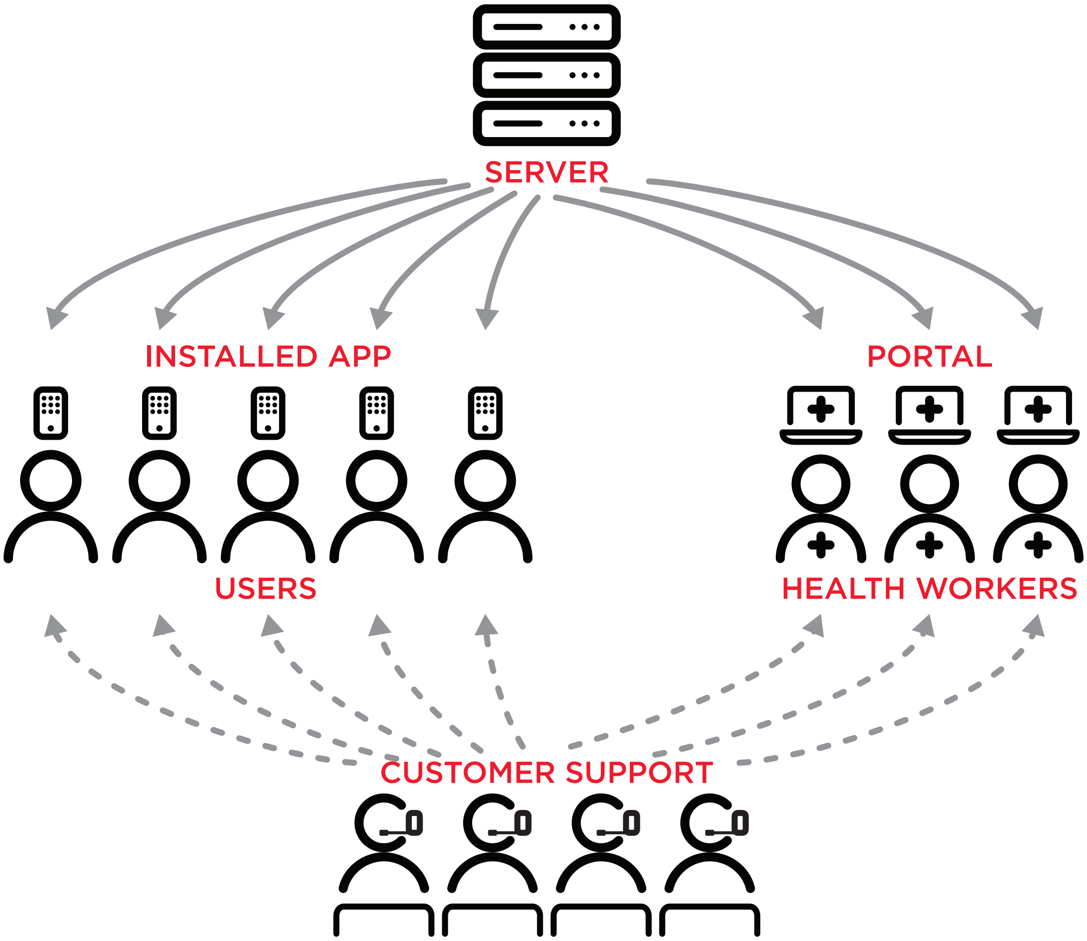
This briefing refers to a contact tracing or exposure notification app, but it is crucial to recognize that a CT/EN solution is, in fact, a platform, the application being but one component of an ecosystem of technologies serving a collective function. A Bluetooth CT/EN platform incorporates the following components:

The app: when two devices come into close proximity, they exchange anonymized IDs, which are then stored in each device's contact log. There should be versions of the app for both iPhone and Android.

A central server: a user who tests positive for COVID-19 may consent to upload an anonymized report—the content of which differs by model—to a central server. This allows the platform to alert other users with whom they may have come into contact, in order to recommend further guidance such as self-reporting, testing, and/or self-isolation. Exposure notification requires only the transfer of anonymized tokens to the central server. Contact tracing requires storage of generalized location information (e.g. half of a postal code) and greater access to personal information. Models employing artificial intelligence (AI) also necessitate vast quantities of personal data, such as health, age, and specific location data, to be stored and/or accessed on a central server.

A health provider web portal connects health authorities to the central server, enabling them to manually authenticate self-reported test results. If centralized contact tracing is employed, the portal also provides health workers with the epidemiological data needed for outbreak tracking.

Customer support: a nationally distributed CT/EN application will take time to be fully functional. Users will likely have questions and concerns regarding the application. Health authorities using the web portal may need support in understanding how to use it. Web developers will need to be made aware of any system bugs and other malfunctions of the platform.



The guiding questions below are categorized into three sections: those for all independent CT/EN proposals regardless of design details; those for Bluetooth CT/EN proposals that incorporate Apple/Google's API; and for those that do not incorporate Apple/Google API. Many security and privacy questions already addressed by Apple/Google's API must likewise be considered when evaluating an independent application.

GUIDING QUESTIONS

SECURITY AND PRIVACY

Due to the highly sensitive use of personal data, as well as the inherent complexity of the platform, addressing security and privacy risks is essential.

The guiding questions below are categorized into three sections: those for all CT/EN proposals regardless of design details; those for Bluetooth CT/EN proposals that incorporate Apple/Google's API; and those for proposals that do not incorporate Apple/Google API. Many security and privacy questions already addressed by Apple/Google's API must likewise be considered when evaluating an independent application.

Questions for all proposals

What are the terms of use for collected personal and health data?

The proposal should dictate who can access, utilize, and share data collected by the platform and for what purpose. This designated team should be limited to health authorities, and with user consent, epidemiologists. Other government agencies, police services, and private companies and developers should be barred from acquiring this data. These terms of use should be communicated to the user.

How long will data be retained by the device? How long will data be retained by the server?

Data should be retained only as long as it is needed for the platform's specific functionality. Exposure notification necessitates that data be stored for 14 days, after which the risk of disease transmission from a given encounter has passed. Contact tracing, which aims to track and monitor outbreaks, and AI-based systems, whose predictions are based on large databases of user information, both retain data long past a two week period. CT proposals, particularly those which utilize AI, must explicitly define a timeline for data retention and deletion that accords with known disease transmission timeframes. Users must be informed of these precise time frames.

Data shared for epidemiological research must be optioned through a clear opt-in and be permanently deleted once shared with epidemiological partners. Any data generated or collected by any component of the platform must be permanently deleted once the system is no longer in use. The apps themselves should be easily deactivated and permanently deleted by users at their own discretion.

What protections are in place to prevent the uploading of false infection reports to the server?

Should fake or erroneous positive test results be accepted, malicious actors could upload their contact logs or tokens (depending on the model used), falsely notifying other users of a potential infection. To prevent these attacks, it is crucial that health authorities manually verify all positive test results. Human verification tests could also be used to ensure that users who report having tested positive are authentic.

What protections are in place to prevent malicious actors from tampering with or surveilling the health authorities' web portal?

Health workers will use the web portal to manually verify positive test results, a process that requires confirming a user's identity. This procedure is arguably the platform's greatest vulnerability. Protections to the web server must be put in place to prevent malicious actors from interfering with or surveilling this process. Moreover, if health authorities maintain a database of infected users, it should be made clear where this data will be stored, for how long, and how it will be protected and encrypted.

What protections are in place to prevent false or multiple app installations?

A way to ensure that all app installations are legitimate may be a human verification measure, like a CAPTCHA test, or an identity verification measure, like requesting a small piece of personal information about the user.

Constant and/or meaningless requests could be deployed to spam the central server and slow down functionality. Security precautions should therefore be taken to ensure that the server is protected from requests that do not come directly from the app.

What protections are in place to prevent a malicious actor from causing a user's app to wrongly notify the user that they have been exposed?

A malicious actor may hack the server or individual devices to wrongly notify a user that one of their contacts has tested positive for COVID-19. If attacking the server, a malicious actor may upload fake contact logs (if a CT model is adopted) or fake tokens (if an EN model is adopted), which could then trigger the distribution of erroneous notifications of potential exposure. A malicious actor could also target individual devices and fabricate notifications that mimic those sent by the application or by government health officials. Security measures should therefore be taken to ensure the authenticity of exposure notifications.

Questions for proposals without Apple/Google's API

What information is uploaded to the server upon diagnosis?

Exposure notification does not require users to upload their contact logs. In fact, EN models explicitly require contact logs to be kept only on individual devices. Uploading contact logs to a central server heightens security risks, as it is more convenient to hack one server than thousands of individual devices. For contact tracing (CT), the platform requires both contact logs and location data (either specific, like an address or GPS coordinates, or generalized, like a general area or half of a postal code). Both specific and generalized location data can be reidentified with relative ease, increasing the risk to personal privacy.

How are tokens generated—by individual devices or by the central server?

Contact logs do not contain personally identifiable information such as names; they instead contain randomized, numerical "IDs" (long lists of randomized numbers), which are generated by daily "tokens". How these tokens are created greatly affects the platform's security risk.

The first option—in which each device generates its own tokens—is widely considered to be the more secure. If contact logs are stored on individual devices, then it is likewise possible for tokens to be individually generated. If a user tests positive, their tokens for the past fourteen days are uploaded to the server. Other devices can then download the infected user's tokens and use them to regenerate the same IDs broadcasted by the infected user. Should a match be found between the infected user's IDs and those in another user's contact log, the other user will be alerted of a potential contact risk. If a malicious actor obtained a contact log, they would also need to obtain the tokens of each user in order to concretely identify users. This would require hacking potentially hundreds of devices.

The second option—in which the central server periodically generates one token for all users—is far less secure. If contact logs are uploaded to the server, it is the server itself that checks each ID and alerts that user's device of a possible infection. To facilitate this, all IDs must be derived from one server-generated token. If a malicious actor obtained a contact log, they would only need the server's tokens to potentially reinterpret the user identities in the log.

Is all data that is collected strictly necessary for the functionality of the platform?

The platform should collect only the amount of data required to achieve functionality and no more. Platforms should strive for minimal data collection and maximum data protection. Collecting location data is unnecessary for exposure notification and increases the risks to user privacy, as this data is usually linked to identifiable information, such as a person's home, workplace, or school. Collecting additional personal information, like age or health data, further heightens those risks, and is unnecessary for both exposure notification and contact tracing.

AI-based platforms require enormous amounts of this personal information to make individual risk predictions, the accuracy of which cannot be absolutely guaranteed. Since massive troves of personal data must be stored on the central server, such models are particularly vulnerable to security breaches, data abuse, and privacy risks. For any alleged gains of AI-based predictions, there is a high-risk tradeoff with the amount of data collected and the perceived and real privacy risks. It is especially important to note that AI solutions are in no way required to deliver either CT or EN platforms.

Could the identity of a user who tested positive for COVID-19 be determined by someone with whom they have come into contact?

Proposals should clearly outline whether additional information, such as time and location of the exposure, would be included in its notification alerts, or if the notification will simply alert that an exposure has occurred. For example, if a notification includes the time at which the exposure occurred, users may infer that who they were with at that time has tested positive, thereby making them privy to private health information.

Questions for proposals with Apple/Google's API

If Apple/Google's model is adopted, all questions from the previous section are already addressed by Apple/Google's API. Apple/Google have developed an exposure notification framework in which individual anonymized tokens are uploaded to a central server. This framework expressly prohibits the collection of location data.

ADOPTION AND IMPLEMENTATION

It is estimated that CT/EN apps need at least 60% participation of the total population to successfully curb the spread of COVID-19.[1] Therefore, the facilitation and encouragement of public adoption should be considered when deciding upon a model.

Will the code used on the platform be open source?

Developers should allow as many components of the platform's code as possible to be open source, including the code for the app, server, and web portal. Apple/Google's API is not open source, so adopting their model would forgo transparency and some associated security and adoptability benefits. However, the Apple/Google API adequately addresses other security questions as outlined above.

Adopting an open source model would allow for technological accountability, as industry professionals and security experts could review the source code and confirm to the public that the system does what it claims to do (and no more). These experts can also publicly address any existing flaws to be remedied. The more people reading the code, the more swiftly bugs and code errors can be identified and fixed. Canadians can be assured that the platform does not have any undisclosed capabilities and works as marketed, thereby assuaging privacy concerns.

Building an open source platform would force developers to integrate and prioritize security at every step of the implementation, rather than relying on security through secrecy. An open source model would also facilitate the modification of the platform to suit regional needs.

Is the proposal scalable for nationwide deployment?

The central server will handle requests from, ideally, millions of devices. Implementation partners should have a proven capacity to build, deploy, and maintain a nationwide high-usage platform. The proposal should address how bugs will be identified and fixed, how new updates will be deployed, and how the platform will be maintained, how often, and by whom. Moreover, the developers should demonstrate the ability to provide widescale technology and customer support to address any user concerns that may arise.

How will public adoption be encouraged and sustained?

Engendering public trust is critical: the app must clearly communicate what data is used, the duration for which it is kept, and how that data is protected in adherence to the Personal Information Protection and Electronic Documents Act (PIPEDA).

A rigorous marketing campaign should be undertaken to encourage maximum public adoption. Aside from publicly ensuring the protection of personal privacy, marketing of the application must be carefully chosen so as to appeal to the widest demographic of people. The implementing partner should provide details about how they will support such efforts.

[1] Hinch, Robert, Will Probert, Anel Nurtay, Michelle Kendall, Chris Wymant, Matthew Hall, Katrina Lythgoe, Ana Bulas Cruz, Lele Zhao, Andrea Stewart, Luca Ferretti, Michael Parker, Ares Meroueh, Bryn Mathias, Scott Stevenson, Daniel Montero, James Warren, Nicole K Mather, Anthony Finkelstein, Lucie Abeler-Dörner, David Bonsall, and Christophe Fraser. 2020. "Effective Configurations of a Digital Contact Tracing App: A report to NHSX". Oxford: University of Oxford.

Certain technologies like AI are negatively perceived by large swaths of the Canadian public. According to a 2018 Proof Inc. study[2], only 25% of Canadian consumers trust AI companies. If an AI-based app is endorsed, this widespread mistrust is likely to induce low adoption rates and extended controversy, eradicating the app's ability to successfully curtail the virus.

How can the application be customized to suit provincial and regional adoptability?

The platform developers and implementing partners must explain, in detail, how the developers would work with provinces to modify the application as needed, including which specific technical aspects would be modified, for how long or how often, and by whom.

RECOMMENDATION:

Given the heightened privacy and security risks inherent to contact tracing and AI-based technology among the available technologies reviewed, this briefing strongly recommends the adoption of an exposure notification model.

[2] Proof Inc. "CanTrust Index 2018", https://www.getproof.com/wp-content/uploads/2020/05/Proof_CanTrust_2018.pdf