Affiliation:

Duke University Bass Connections
The Privacy Implications of COVID-19 Contact Tracing
Technology Team

Authors:

Ana DeCesare, ana.decesare@duke.edu
Phoebe Dijour, phoebe.dijour@duke.edu
Leah Markbreiter, leah.markbreiter@duke.edu
Jerry Xin, mx47@duke.edu
Jessica Yang, jessica.yang@duke.edu

Trade-Off Between Privacy and Efficacy

As a response to the global pandemic, government and private entities have increasingly turned to technology in the hopes that harnessing it properly will help subdue COVID-19 more quickly and effectively. A significant portion of these efforts include the development of mobile applications that at least support manual contact tracing. These contact tracing apps work under the assumption that if the population uses them properly, then outbreaks can be tracked faster, and public health authorities can more quickly act accordingly.

Contact tracing, though proven to be an effective method of disease control when done manually, by nature requires members of the population to give up information about their lives. In fact, the more information that can be shared by positive patients about their daily activities and who they may have come into contact with, the more effective the contact tracing is. As a result, contact tracing apps that collect information on their users in the name of public health may do so at the cost to the user's privacy. This is especially true as digital apps installed in a user's phone which will have access to significantly more personal information than would be given up by a COVID positive patient to a human contact tracer, and at a greater granularity of detail. However, because the efficacy of contact tracing is influenced by the amount of data that can be provided, failing to collect enough relevant data in a bid to protect user privacy would defeat the purpose of the app entirely.

Further increasing the difficulty of this balancing act between privacy and efficacy is the impact of privacy intrusions and security breaches on user trust and compliance. Both manual contact tracing and contact tracing apps rely on the participation of the population in order for it to be effective. Users may lose trust in an application's ability or willingness to properly handle and secure sensitive personal information, and therefore result in decreased use of the application. At that stage, an application will have mishandled user data entrusted to them, only to fail in the stated purpose of the application.

In the interests of better understanding what circumstances are required for contact tracing apps to be effective, we will model an infection spreading through a population under different scenarios involving contact tracing apps. In particular, we will be comparing different levels of app compliance, scenarios in which apps have been attacked through security vulnerabilities, and reactive versus proactive contact tracing. Through this, we hope to better understand the conditions necessary for a successful app, and what the fallout would be in cases where contact tracing apps fail.

Agent-Based Threat Models

In order to simulate some of the anticipated privacy attacks with digital contact tracing, we will create two agent-based threat models. The first will feature the centralized approach to contact tracing: a network of mobile devices with user data stored on a central server. The second will demonstrate the decentralized approach: a network of mobile devices with user data stored only on a local device level. For each model, we will first show the standard and expected outcome of digital contact tracing, namely the safe and reliable spread of necessary COVID-19 information through the network. Then, we will simulate a variety of privacy attacks, including linkage, false positive, relay, nerd, militia, and paparazzi attacks, as well as unintended consequences, such as over-inclusion.

In order to simulate these scenarios, several assumptions must be applied. For all simulations, we will assume a certain population adoption percentage and user compliance percentage with assumed app guidelines. Each attack will additionally have its own assumptions, which will be discussed further. Quantitative results of user data leakage will be averaged over several hundred random trials and tabulated.

Agent-Based Infection Spread Model

Given the data-driven hypothesis that a greater number of privacy intrusions will decrease public trust and thus app adoption rate, we will generate an agent-based model of viral spread through a network of app users. This model, separate from the previously mentioned agent-based threat models, will not demonstrate the spread of information through a network of devices, but rather the spread of infection through a network of individuals based only on app adoption and compliance. Several parameters, including size of population, average degree of interactions, initial outbreak size, percent of infected individuals that are asymptomatic, infection rate, incubation time, boolean ability to spread the virus during incubation period, recovery time, and chance of death, can all be adjusted in the user-interface. Each of these parameters has a set of assumptions that will be discussed further. For instance, each symptomatic individual quarantines after diagnosis, while each asymptomatic individual does not. We will first simulate this infection model

Moreover, in this model, both reactive and preventative contact tracing approaches can be demonstrated, and percent app adoption for each approach can be chosen on a sliding scale from 0 to 100%. The reactive approach ensures that each individual has a certain probability, dependent on app adoption rate, of quarantining when a primary contact has tested positive. The preventative approach ensures that each individual has a certain probability, dependent on app adoption rate, of isolating when the virus is somewhere in his/her network of normal interactions, regardless of if they have actually interacted. In the preventative approach, individuals can avoid exposure to the virus in their network, even before interacting with a COVID-positive individual. We will simulate the spread of infection with no contact tracing and with different adoption percentages for each of the two contact tracing approaches. We will average and tabulate maximum infection and total death data and compare the results to simulate what percent adoption is necessary for a desired app efficacy.