Peer to Peer Contact Tracing: A Smartphone Application for Tracking Routes of Transmission without Collecting Personal Data

Abstract

Contact tracing using smartphone technology is a powerful tool which may be employed to limit disease transmission during an epidemic, yet contact tracing applications present with significant privacy concerns regarding the collection of personal data such as location. We propose a novel, peer to peer mechanism for tracing potential routes of disease transmission without the collection of personal data and develop a smartphone application prototype which implements this method. Additionally, we develop a computer simulation which demonstrates the impact of our proposal on outbreak trajectories across multiple rates of adoption.

Introduction

The COVID-19 epidemic represents an urgent public health crisis. According to epidemiologic modeling conducted by Ferguson *et al*., both the US and the UK face a dilemma in choosing a public health response.[1] On one hand, the models predict severe consequences, including high death rates, if the virus is permitted to run its course without any intervention or response. However, the authors conclude that an optimal outcome following a strategy of disease suppression would likely require dramatic alterations to daily life, including social distancing, for the entire population until a vaccine is available. Such an intervention may result in significant economic loss. Ferguson and colleagues noted that technologic solutions, such as a contact tracing application, might provide alternatives to the drastic measures they proposed.

However, such an application presents significant concerns regarding privacy. Recent events, such as the Equifax security breach or the Cambridge Analytica controversy regarding data collection through Facebook, have highlighted these concerns, which are especially pertinent in a health care setting.[2-4] Most existing contact tracing applications rely on the collection of personal data, such as location, in order to determine exposure risk. Privacy concerns have already been raised over government-sponsored contact tracing applications in Iran, South Korea, and China.[5]

While the balance between privacy and other objectives is certainly controversial, there is an additional concern in the case of a contact tracing application. As we will demonstrate shortly, the efficacy of a contact tracing application depends on its adoption rate. Even those that may not prioritize privacy over other concerns should be concerned about the efficacy of a contact tracing application that is believed by many to be an invasion of privacy. If a sufficiently large portion of a population does not participate due to privacy concerns, such an intervention may produce limited results.

A Novel Approach

Contact tracing is the process of tracking routes of transmission of an infection through a population for the purposes of isolating those who may have been exposed and reducing further transmission. Contact tracing, in varying forms, has been utilized for a number of diseases including tuberculosis and HIV.[6] While location-based contact tracing presents a viable solution, it is not without limitations. Location data is highly personal, and the privacy concerns detailed above are especially salient for location data.[7] Furthermore, location is only a proxy for contact, and inferences about exposure based on location may not always be accurate due to noise in the data.[8]

Here, we propose a novel method for contact tracing using a smartphone application without the use of any location information. The objective of this method is to provide an effective contact tracing mechanism without compromising user privacy.

At the core of this method is a data structure which we will call the *transmission graph*. The transmission graph consists of nodes, which represent *contact points* between individuals, and directed edges, which represent *transmission vectors* between contact points. Whenever an individual participates in a contact point, a transmission vector is added to the transmission graph from the individual's prior contact point to the current contact point. The transmission graph, then, is a network of interactions between individuals. Of note, there are no entities in the graph which represent a particular individual, and location information is never encoded in this data structure. These properties of the transmission graph are fundamental to its privacy-preserving nature.

Each contact point (node) in the transmission graph can be in one of two states: status positive (SP) or status negative (SN). A SP contact point is one which has been flagged as having one or more participating individuals which was positive for infection at the time of interaction. A SN contact point is simply one that has not been marked SP. SP and SN are merely internal states of the graph. As will be demonstrated shortly, they are distinct from the actual states which will be displayed to end users.

Using the simple data structure of the transmission graph, *possible transmission paths* can be determined for any given target contact point. A possible transmission path is defined as a path from a SP node to a given target node. For any given target node, there may be 0, 1, or multiple possible transmission paths.

Smartphone Application

We have developed a prototype of a smartphone application utilizing the transmission graph data structure to implement peer to peer contact tracing. The application code is open source and publicly available at https://github.com/tyleryasaka/covid-watch. In the current prototype, users are assigned one of two risk levels: standard or elevated. The user is considered to have elevated risk level if they have any recent contact points with possible transmission paths, and standard risk level otherwise. Here, "recent" refers to some predetermined amount of time after which an individual is assumed to have recovered (*infectious period*). Based on available data, we have tentatively set this time to 21 days.[9]

When a user receives and reports a diagnosis, it is assumed that the infection was present for some time prior to the diagnosis. The application will determine the earliest contact point which occurred before this time and it to the server accordingly. The amount of time that a user is assumed to have been infected is another parameter (*diagnosis delay*) that can be adjusted according to data and expert opinion. It is currently set to 14 days to follow the quarantine protocol recommended by the WHO.[10]

For the purposes of the smartphone application we are proposing, a key concept is the storage of and access to information. In the current prototype, all transmission graph data as described above is stored on a centrally managed server. Importantly, no user registration is required and no personal information is collected; thus, the data on the server is anonymized. User smartphone applications will only be able to access the user's own checkpoints, and whether there are any possible transmission paths to those checkpoints. An illustration of a hypothetical disease spread scenario, along with the corresponding transmission graph and the information available to the server and each user application, is provided in **Figure 1**.

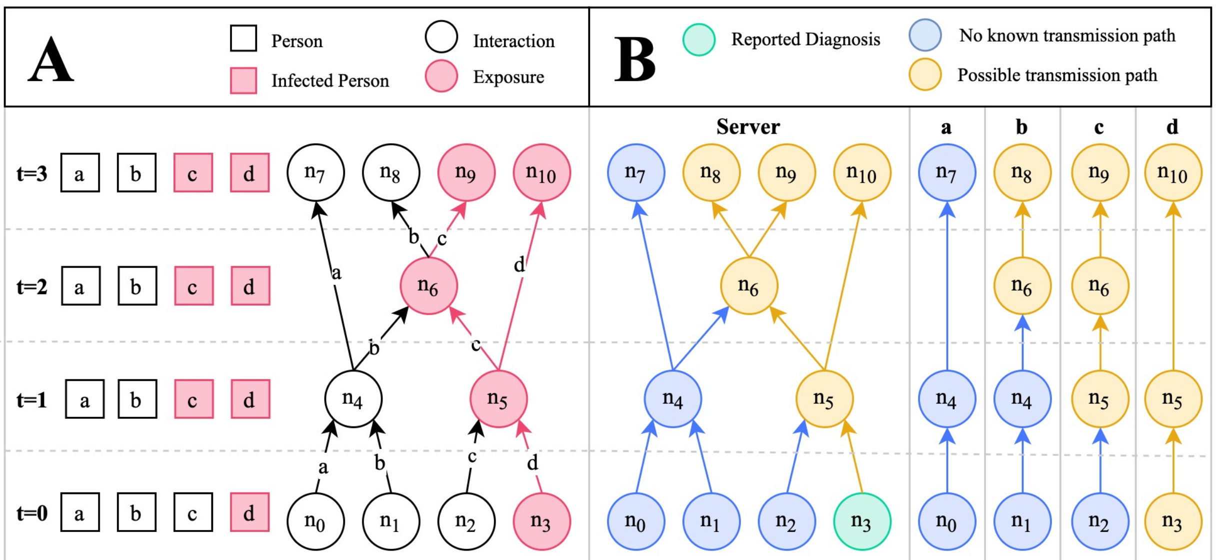

Figure 1: Disease spread scenario modeled as a transmission graph
A. Graphical representation of a disease spread scenario across time. Nodes represent contact points, while edges represent the paths of individuals from one contact point to the next. Contact points with infected individuals are marked as exposures. Uninfected individuals may become infected at exposure points according to some probability; hence, *b* does not become infected at n_6. **B.** Transmission graph corresponding to the scenario in *A*, depicting the information that is available to the server and each individual's smartphone application. Only one node, n_3, is associated with a reported diagnosis. The infection risk level at the other contact points can be inferred by checking for possible transmission paths.

Another important characteristic of the proposed application is its peer to peer nature. Traditional contact tracing applications often rely on a central entity to monitor individuals, their infection status, and their locations.[5] Our proposal does not require or utilize such entities. Instead, individuals rely on the joint participation of their peers, both in terms of creating contact points and reporting diagnoses. No external entity is required to monitor the central server or database, other than to ensure its proper functioning.

In the current application prototype, contact points are created by one person becoming a "host"; this displays a QR code on the user's screen. Other users can then join the contact point by scanning the QR code. Additionally, users must press a button in the application in order to report a diagnosis. At any time, users may open the application to check their current risk level.

Simulation of Peer to Peer Contact Tracing

Rigorous mathematical models of contact tracing have been previously described in the literature, and we do not attempt to replicate them here.[11,12] Rather, we develop a low-fidelity computer simulation model which facilitates disease spread through interaction of individuals at contact points across time, allowing for the explicit modeling of the transmission graph structure we have proposed here. With this model of disease spread, we are able to compare outbreak trajectories both with and without peer to peer contact tracing as we have described. Such a model, while not intended to describe real-world trajectories, allows for demonstration of the feasibility of our proposal, and provides a rudimentary mechanism to compare various scenarios and application parameters, such as the adoption rate, the diagnosis delay estimated by the application, and the infectious period estimated by the application. The model source code is publicly accessible at https://github.com/tyleryasaka/covid-watch and a public, web-based interface is provided at https://tyleryasaka.shinyapps.io/covidwatch. A comparison of simulations produced by this model with varying levels of adoption is presented in **Figure 2**.

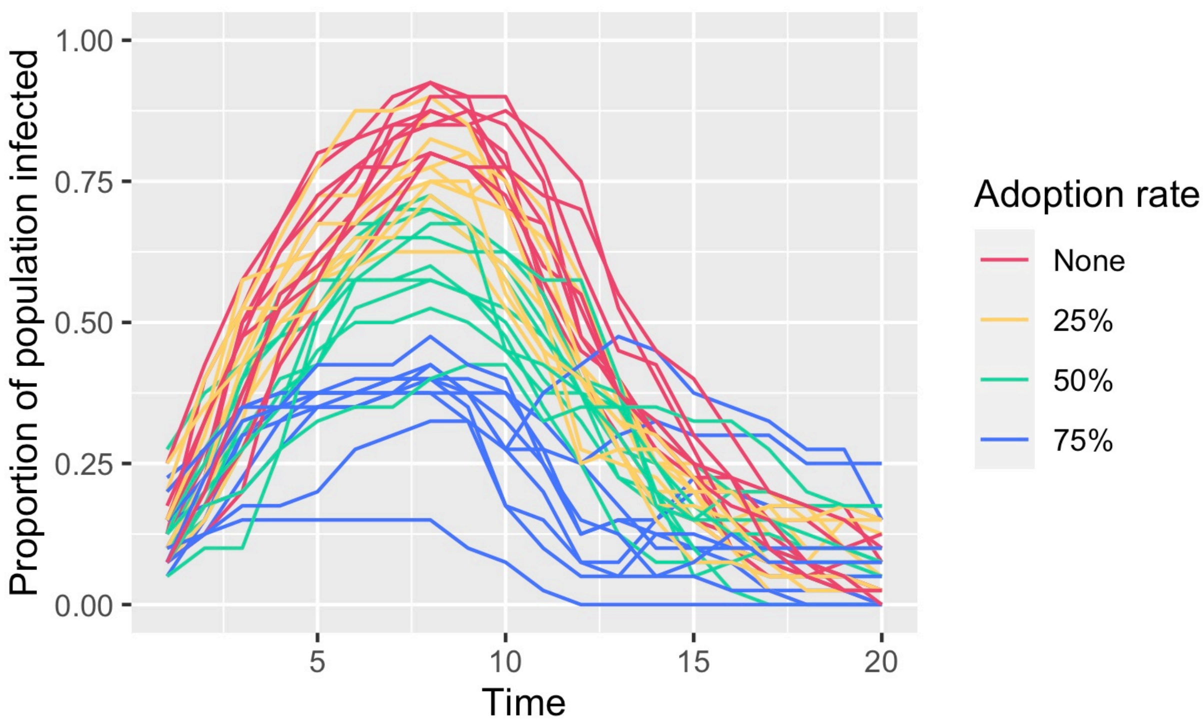

Figure 2: Comparison of infection curves from simulations at varying rates of peer to peer contact tracing application adoption
The proportion of the population with active infection is plotted across time for multiple adoption rates: 0% (none), 25%, 50%, and 75%. The results of 10 random simulations per adoption rate are given.

Privacy and User Adoption

User adoption is a critical factor in the success of any contact tracing application. Our simulation results suggest that if participation is not high, the infection will continue to spread through the population via those not using the application and the intervention will have limited effect. We have discussed privacy concerns already; it is worth emphasizing that in certain countries (such as the US), privacy concerns may present a significant barrier to adoption.[13] Location tracking, in particular, is a type of data collection that may deter many from using an application,[14] especially when this data is being collected by or shared with government entities.[13] By not requesting access to location data, our application avoids this potential hurdle to adoption.

Development Time Frame and Complexity

Two aspects of developing smartphone applications that are often overlooked are development time and complexity. Complex software not only takes longer to develop, but is also more prone to failures; a noteworthy example was the failed initial launch of the HealthCare.gov website.[15] The COVID-19 pandemic represents a time-sensitive public health crisis, and any technologic solution applied to this crisis would need to be developed in a manner that is not only rapid but is robust to failure. The smartphone application we have proposed is highly simplistic in its design and could be deployed swiftly without sacrificing robustness. Additionally, while our proposed application would ideally be released under the endorsement and/or guidance of government entities, it would not require the type of intimate data sharing and coordination that are implied by alternative designs. Thus, bureaucratic overhead would be minimal.

Conclusion

We have proposed a novel peer to peer smartphone application for contact tracing which does not utilize personal data such as location, and hence preserves user privacy and anonymity. Such an application could potentially be applied to the COVID-19 epidemic as well as others in the future, in order to achieve a middle ground between drastic isolation measures and unmitigated disease spread. We have developed a prototype of this application which is open source and publicly available, as well as a computer simulation model which demonstrates the potential of our application to impact the course of an epidemic.

References

1. Ferguson NM, Laydon D, Nedjati-Gilani G, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *London: Imperial College COVID-19 Response Team, March.* 2020;16.
2. Isaak J, Hanna MJ. User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer.* 2018;51(8):56-59.
3. Iyengar A, Kundu A, Pallis G. Healthcare informatics and privacy. *IEEE Internet Computing.* 2018;22(2):29-31.
4. Gostin LO, Halabi SF, Wilson K. Health data and privacy in the digital era. *Jama.* 2018;320(3):233-234.
5. Raskar R, Schunemann I, Barbar R, et al. Apps Gone Rogue: Maintaining Personal Privacy in an Epidemic. *arXiv preprint arXiv:200308567.* 2020.
6. Armbruster B, Brandeau ML. Contact tracing to control infectious disease: when enough is enough. *Health care management science.* 2007;10(4):341-355.
7. Smith M, Szongott C, Henne B, Von Voigt G. Big data privacy issues in public social media. Paper presented at: 2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST)2012.
8. Farrahi K, Emonet R, Cebrian M. Epidemic contact tracing via communication traces. *PloS one.* 2014;9(5).
9. Bi Q, Wu Y, Mei S, et al. Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. *medRxiv.* 2020.
10. Organization WH. *Considerations for quarantine of individuals in the context of containment for coronavirus disease (COVID-19): interim guidance, 19 March 2020.* World Health Organization;2020.
11. Kwok KO, Tang A, Wei VW, Park WH, Yeoh EK, Riley S. Epidemic Models of Contact Tracing: Systematic Review of Transmission Studies of Severe Acute Respiratory Syndrome and Middle East Respiratory Syndrome. *Computational and structural biotechnology journal.* 2019.
12. Browne C, Gulbudak H, Webb G. Modeling contact tracing in outbreaks with application to Ebola. *Journal of theoretical biology.* 2015;384:33-49.
13. Dinev T, Hart P, Mullen MR. Internet privacy concerns and beliefs about government surveillance–An empirical investigation. *The Journal of Strategic Information Systems.* 2008;17(3):214-233.
14. Xu H, Gupta S. The effects of privacy concerns and personal innovativeness on potential and experienced customers’ adoption of location-based services. *Electronic Markets.* 2009;19(2-3):137-149.
15. Borchers AS, Crawford J. Healthcare. gov Website Failure. *Journal of Critical Incidents.* 2014;7.