The Limits of Location Tracking in an Epidemic

By Jay Stanley and Jennifer Stisa Granick
April 8, 2020

As Americans struggle to confront the COVID-19 outbreak, some have suggested that cell phone location tracking technology can help in the effort to contain the disease. The tech industry and the White House are reportedly having conversations over how information technology might be deployed, and there is increasing discussion about how foreign countries are using technology. The governor of Florida has even floated the idea of using an app to track visitors from COVID-19 hotspot New York.

However, policymakers must have a realistic understanding of what data produced by individuals' mobile phones can and cannot do. As always, there is a danger that simplistic understandings of how technology works will lead to investments that do little good, or are actually counterproductive, and that invade privacy without producing commensurate benefits.

As we write this white paper, public health experts say that the nation has three urgent needs: strong social distancing measures, widespread testing capability, and material support for hospitals being overwhelmed by victims. However, once our hospitals reach a point where they're able to handle the stream of new patients, experts say that indiscriminate population-wide social distancing measures may give way to a new phase: chronic, lower-level waves of infection in which a combination of widespread testing, individualized quarantine orders, and traditional epidemiological contact tracing once again become a principal means of combatting the disease. It would be in such a period — the window between the end of the initial wave and the development of a vaccine — that using certain forms of data generated by cell phones — such as location histories or records of proximity to other devices — might make sense.

The challenges posed by COVID-19 are extraordinary, and we should consider with an open mind any and all measures that might help contain the virus consistent with our fundamental principles. We note some of those possible uses in this paper. At the same time, location data contains an enormously invasive and personal set of information about each of us, with the potential to reveal such things as people's social, sexual, religious, and political associations. The potential for invasions of privacy, abuse, and stigmatization is enormous. Any uses of such data should be temporary, restricted to public health agencies and purposes, and should make

the greatest possible use of available techniques that allow for privacy and anonymity to be protected, even as the data is used.

Because location tracking has such ominous potential implications, in short, we should make sure that any uses of such sensitive data are necessary, effective, and proportionate. The principal focus of this paper is on whether such uses would be effective.

Key Questions

The likely effectiveness of different uses of cell phone location data hinges on exactly how each is envisioned as being used in the effort to stem the spread of the coronavirus. Important variables are:

1. **What is the goal?** Is it tracking overall trends, helping people who have tested positive recall past contacts, identifying unknown individuals who may have been infected by the patient, or enforcement of quarantines or stay-at-home orders?
2. **What data?** Is it aggregate and anonymized data, or individually identifying information? How precisely can the information pinpoint individuals' locations? Is the dataset complete enough that one can draw meaningful conclusions? Will the data under- or mis-represent people of color or low-income communities in a manner that could lead to prejudicial results, such as inferior access to health care or over-policing?
3. **Who gets the data?** Does the government get access to the raw data, is it shared only with public health entities such as qualified academics or hospitals, or does it remain in the hands of the private entity that originally collected it?
4. **How is the data used?** Is it used for centralized government action, such as issuing or enforcing quarantine orders, or for punitive measures? Or, does it enable decentralized individual decision-making such as checking announcements of possible exposure points and choosing to go to a testing center?
5. **What is the life cycle of the data?** Any corpus of data is likely to create risks to the people it represents. A responsible steward of other peoples' data will have plans for data destruction once the data's relevance is diminished, to mitigate future compromise.

The answers to these key questions vary widely for the different proposals that have been the center of recent discussions. Now, we turn to each of these proposals and consider them in light of the key questions.

Proposed Use: The use of mass location data to identify unknown individuals who may have been exposed to a contagious person

One often-mentioned idea is to gather mass location data on as many people as possible and analyze that data with the goal of identifying those who have been exposed to an infected person. Under this vision, the location record of a newly discovered coronavirus-positive person would be matched against everyone else's location records, and those who were near the infected person would be notified, tested, and/or ordered into quarantine. This kind of tracking is part of what China has been doing in its seemingly successful effort to suppress the virus,

which has fed the appeal of such tracking. Israel has also revealed that its government has been secretly retaining detailed cell phone location records on its population, and has begun trying to use them in this way against COVID-19. And "about a dozen" other countries are reportedly testing the same concept through a product created by a notorious spyware company.

There are serious practical problems with this concept, however.

Data on individuals' locations is not accurate enough for automated contact tracing.
We have spoken with engineers and executives at a number of the largest U.S. companies that hold location data on Americans' movements and locations, and generally they have told us that their data is not suitable for determining who was in contact with whom for purposes of COVID-19 contact tracing.

Data on Americans' movements and locations is collected through various technologies:

- Cell tower location data**.** Cellular phones continuously emit signals to identify themselves to nearby towers or other cell sites so that the system can route data and calls their way. Each time a phone connects to a cell site, it generates a time-stamped record known as cell-site location information (CSLI). Wireless carriers collect and store this information. Because the carrier knows the physical location of the cell tower, this data has the side effect of telling telecom providers approximately where those phones are. The precision of cell-site location information varies depending on the strength of the signal, the density of towers, network load and other factors. Numerous data entries over time can support an inference that a phone was near a particular location at a certain point in time. They can also show when a phone is moving in a particular direction — for example, southbound on highway 101. But the data is not precise enough to tell you how close two phones are to each other. In rural areas the information may place people only in a several-square-mile area. In urban areas, where towers are more densely deployed, the accuracy is typically higher, but still insufficient for reliably pinpointing location. China reportedly looked into using cell-site location data in its fight against COVID-19 but found that it was too inaccurate, generated too many false positives, and was wasting manpower.
- GPS. When mobile phones have satellite sensing turned on, they can identify their own location with a best-case theoretical accuracy of 1 meter, but more typically 5 to 20 meters under an open sky. GPS radio signals are relatively weak; the technology does not work indoors and works poorly near large buildings, in large cities, and during thunderstorms, snowstorms, and other bad weather. It can also take a GPS receiver up to several minutes to perceive its location when first turned on or brought outdoors. In addition, phones *receive* transmissions from GPS satellites, but they do not automatically calculate, save, or broadcast GPS location information. That only happens when a user installs an app on their phone that is designed to do that. Those apps may send that data to dozens of different companies or to none at all; there is not a single repository.

- Wi-Fi, Bluetooth, and other radio location. Google and other companies hold databases identifying Wi-Fi router locations, and use that information in combination with GPS data to help locate phones with more precision than can be done with GPS alone. Bluetooth, the wireless technology used to exchange data over short distances (less than 30 feet), is also widely used for tracking, especially in stores. But not all phones have Bluetooth turned on by default and Bluetooth contacts do not comprise anything like a comprehensive tracking system.
- QR codes. In China, residents are required to install an app on their phone that they must use to scan QR codes that have been placed in taxis and at the entrances of buildings, buses, and subway stations. These real-world checkpoints provide far more reliable and accurate tracking than wireless technologies — and can be combined with those technologies. The United States, however, has no such checkpoint infrastructure in place, little capacity or apparent desire to build one, and no legal authority to compel people to carry a phone, much less install a specific app on their phone.

The CDC warns against coming in "close contact" with a person who has tested positive for the virus, defining close contact as "being within approximately 6 feet (2 meters), of a person with COVID-19 for a prolonged period of time." None of the data sources discussed above are accurate enough to identify close contact with sufficient reliability. None are reliably accurate to within 6 feet. Using the wrong technology to draw conclusions about who may have become infected might lead to expensive mistakes such as two week isolation from work, friends, and family for someone — perhaps even a health care worker or first responder — who was actually not exposed. Israel's use of location data has already sparked complaints about accuracy.

A location tracking system over time can be accurate enough to place a person near a bank, bar, mosque, clinic, or other privacy-sensitive location. But the fact is, commercial location databases are compiled for advertising and other purposes and are simply not accurate enough to reliably determine who was in close contact with whom.

The algorithms are not likely to be reliable.

Even if we were to imagine a set of location data that had pinpoint accuracy, there would still be problems translating that in any automated way into reliable guesses about whether two people were in danger of transmitting an infection. The Israeli system apparently acts on the basis of nothing more than an automated look at proximity. In Israel, one woman was identified as a "contact" simply because she waved at her infected boyfriend from outside his apartment building — and was issued a quarantine order based on that alone. Such a system is likely to make many such mistakes; it won't know that a bank teller is shielded from transmission because they're behind plexiglass, or that two people close together in a building are actually in separate apartments divided by a wall.

The alternative is to try to make more educated guesses by taking account of such circumstances as well as such factors as the duration of a contact and the number of days the positive-testing person has been infected. But those guesses will inevitably be highly unreliable.

Ambitious plans for automated recording and analysis of human life are everywhere these days, but few such efforts are reliable because human life is messy, complicated, and full of anomalies.

The data is fragmented and may be biased.
Complicating matters further, no single, centralized party holds the location data generated by Americans. Most location data with any level of precision is generated by an essentially corrupt ecosystem of shady, privacy-invading companies that engage in mass location tracking without individuals' meaningful awareness or consent, typically by paying the developers of smartphone apps to hide tracking capabilities inside those apps. This location data is scattered among dozens of such companies most Americans have never heard of. And giants like Google and Facebook that collect location data have enormous reach within the population, but only have location data on a minority of their users. Any kind of automated contact tracing that hopes to find close contacts will need to access more than a slice of existing data pools if the tracking is to effectively find otherwise unknown infected people.

There may also be differences in how various populations and demographics are represented in the data. Making public health decisions on such datasets could leave out entire populations, and misrepresent others, and lead to a deployment of health care resources that is not only biased and unjust — tilted toward wealthy neighborhoods, for example — but ineffective from a public health standpoint.

Proposed Used: The use of specific location data to assist in identifying where an infected person was in the recent past

Location data is more effective when used to trigger a patient's recollection of places traveled in the past. South Korea has so far succeeded in minimizing its outbreak through a massive testing and contact-tracing program without the kinds of lockdowns and business closures that other countries have resorted to. One element of its contact-tracing effort is retracing the steps of newly identified coronavirus carriers. To help do that, the legislature has passed a law giving the government authority to collect location data from those who test positive. But instead of trying to cross-reference the location histories of infected people against massive population-wide location datasets, South Korean officials simply "anonymize" and publish the patients' location histories. Numerous web sites and apps make that information available, and citizens can check for possible collisions with infected people. Widespread and quick testing is available for anyone worried that they have been exposed. Such checks have become a part of daily life in South Korea.

Similar mechanisms have been part of the solution in China. Baidu (a Chinese analogue to Google) created a map layer on top of its standard mapping app that shows real-time locations of confirmed and suspected cases of the virus so that people can avoid more-affected areas. Another company, Qihoo 360, launched a platform where travelers can check if anyone on their recent train or plane trips has since tested positive.

One problem that has emerged in South Korea is that the authorities are not doing a good job anonymizing the data. One alert informed the public, for example, of a "43-year-old man, resident of Nowon district" who was "at his work in Mapo district attending a sexual harassment class." Unsurprisingly some people described in such ways are being identified by members of the public. That has reportedly left many people more afraid of social humiliation than of the disease itself, threatening to interfere with testing, arguably the most important intervention to control the virus.

In fact, data scientists have learned in recent years that it is surprisingly difficult to anonymize data. That is doubly true with regard to location data, because people reveal themselves by where they go; even relatively rough information about a person's location will often identify them uniquely. For example, according to one study, just knowing the zip code (actually census tract, which is basically equivalent) of where you work and where you live will uniquely identify 5 percent of the population. If you know the "census blocks" where somebody works and lives (an area roughly the size of a block in a city, but larger in rural areas), the accuracy is much higher, with at least half the population being uniquely identified. And of course, once a person's home address is identified, little doubt as to their identity remains.

Even without robust unique identification, stigma and harassment could follow. Imagine four different people who all live in ZIP code A and work in ZIP code B: Even if they are not uniquely identified, if one tests positive and their location data is published, the other three may fall under suspicion as well.

That said, information about the times and places where infected people were present could certainly be aggregated and anonymized far more effectively than South Korea has done. And there are ways to use location data that don't involve publishing it — for example, it could simply be used in cooperation with those who are infected to jog their memories and help them retrace their movements.

Proposed Use: The use of aggregate location data

The use of anonymous aggregated location data could help epidemiologists answer questions such as how many people are moving between adjacent neighborhoods, towns, or states each day, or how much the average resident of an area changes their travel after new movement limitations are imposed. The CDC and state and local governments in the U.S. are reportedly receiving some such data from the mobile advertising industry.

Aggregate location data does not usually reveal individuals' private information (though there can be exceptions, such as in sparsely populated areas where the numbers are so small that they can reveal an identifiable person's movements). Of course, some party needs to hold the raw data before it is aggregated. Though such data could theoretically be generated using privacy-protecting software built into the collection mechanisms, few players in this ecosystem have the incentive to utilize such mechanisms. So far, the sources of underlying location data in

the aggregate applications we've seen are unsavory — this is data surreptitiously collected via mobile ads and funneled through a location data broker.

One question is who sees what. Ideally, nobody sees anything more than they need. For example, a public health official who needs to know how many people travel daily to her town from a nearby COVID-19 hotspot does not need access to her fellow residents' personalized travel data to find that out. In many cases, the entity that already holds the location data can keep it and the government need not receive a copy. Any entity should be transparent and open about where the data came from, who holds it, and how it is being analyzed.

Another question is what officials do with the insights that even anonymous and aggregated data can generate. Encouraging voluntary social distancing is one thing; increasing police interventions in non-compliant zip codes is another. Will the data create misleading pictures of behavior within communities of color or low-income communities in a manner that could lead to prejudicial results, such as inferior access to health care or over-policing?

Proposed Use: Enforcement

Another possible use of location information is in the enforcement of quarantines, shelter-in-place orders, and travel restrictions. These uses contemplate turning people's cell phones into ersatz ankle monitors.

As a technological matter, the accuracy of cell phone location data is adequate for detecting people who move any significant distance from their homes, or who are traveling from another place.

Aggregate data could also be used to enforce compliance with stay-at-home orders. For example, companies could notify the authorities when cell phone data suggested that people were gathering in numbers prohibited by local public health measures, without sharing the identities of those who appear to have gathered, such as at a rogue bar that appeared in Los Angeles during a shelter-in-place regime. Such measures would have to be used with care lest multi-story apartment buildings or the like trigger alarms.

Public health experts caution, however, that a law enforcement approach to combatting disease is less effective than relying on voluntary measures and compliance. That is because an enforcement approach often sparks counterproductive resistance and evasion and tends to sour the relationship between citizens and their government at a time when trust is of paramount importance. Good public health measures leverage people's own incentives to report disease and help stop its spread.

Indeed, where people feel that their phones have become an instrument of an antagonistic government, there are a number of technological steps they might take to defeat that tool. The simplest step would be to turn the location function off on their phone, or turn their phone off

entirely. Uncooperative subjects could also simply leave their phones at home, or acquire and register a second, dummy phone that is not their primary device which they leave at home.

It is probably to forestall such stratagems that Hong Kong is issuing electronic tracker wristbands to people under compulsory home quarantine to ensure they do not go out. In Taiwan, an American University student under quarantine received a visit from police after his battery died while he slept. To be effective, that kind of close monitoring of cell signals at scale would require either significant manpower and resources or the creation and implementation of automated detection capabilities, and the administrative structures required to back them up — all well before the arrival of a vaccine rendered them superfluous.

Other Possible Concepts and Models

There may be many other ways of using cell phone data that have not yet been fully explored. For example, in Singapore, an app called TraceTogether uses short-range Bluetooth signals to keep track of which other people come nearby over time. If a user later turns out to be positive, their phone then holds an encrypted record of who else was nearby, which can be unlocked by Singapore's Ministry of Health. Rather than relying on location tracking, they are instead using "proximity tracking."

There are a number of problems with the concept. In Singapore, the Ministry of Health has access to user data about every app user, and detailed access to activity and contact logs and from every infected user. While they've promised to use this access only for public health purposes, there is nothing stopping them from turning over data to law enforcement or other punitive agencies. Also, for the app to be effective in its public health goals, it would need a certain level of adoption by the public, which might be averse to installing it out of fear of these punitive measures. Other open questions include whether Bluetooth is precise enough to distinguish close contacts given that its range, while typically around 10 meters, can in theory reach up to 400 meters, and that its signal strength varies widely by chipset, battery, and antenna design.

It may be possible, however, to design implementations that address some or all of these problems and adhere to important principles such as voluntariness, decentralization, simplicity, transparency, and the lack of reliance on a persistent identifier. Such implementations would enjoy increased adoption and effectiveness because they would engender public trust. A number of developers are working on such implementations with the welcome goal of being both privacy-preserving and effective.

And of course, there may be other ideas for leveraging technology that no one has yet conceived.

Conclusion

In this crisis, we need to seriously consider how technology might help improve public health. This investigation must be based on a realistic understanding of what technology and data can and cannot do, lest we divert attention, expertise, and financial resources from other, simpler but time-tested methods that are more effective. In particular, policymakers should understand the limits of existing location data and devices for automated contact tracing.