COVID-19 Geolocation: Leveraging existing technologies and crowdsourcing for exposure risk assessment

Eric Schuldenfrei[A], Ashley Scott Kelly[A], and Howard Huang[B]
[A] The University of Hong Kong
[B] Nokia Bell Labs

31 March 2020

COVID-19 is primarily spread two ways: 1) Via respiratory droplets; and 2) Transfer from a contaminated surface to the face via your hands (WHO, 2020). Awareness of both colocation and whether an individual has contacted a contaminated surface, such as a doorknob, is important and could better guide individuals toward self-quarantine and COVID-19 testing. Therefore, to contain the periodic spread of COVID-19 in communities, it is vital for individuals to know their personal 14-day exposure risk, which is a combination of having crossed paths with a confirmed COVID-19 case and having contacted a potentially contaminated surface.

Our proposal is a smartphone application and exposure risk assessment model that leverages existing technologies supplemented with the crowdsourced data outlined in this paper. The COVID-19 Geolocation App[1] (the "App") records an individual user's location history and computes their exposure risk by cross-referencing that history with an *Infectious Space-Time Map* (ISTM). Exposure risk is computed entirely on one's personal smartphone using a geographical subset of the ISTM, which is updated daily from a central server. If heightened exposure risk is detected, the App displays a notification on the user's smartphone that suggests further action, such as self-quarantine, based on current epidemiological understanding. The ISTM is our proposed model that synthesizes the 14-day[2] location history of voluntarily disclosed (and, in many contexts, health-authority confirmed) COVID-19 cases with existing outdoor and indoor geolocation technologies in public and semi-public spaces. The ISTM focuses explicitly on rapid deployment, user privacy, and flexible adoption of new epidemiological knowledge, such as increased risk due to prolonged exposure to a potentially infected user's symptomatic phase, and geolocation technologies as they become available. If an individual is confirmed with having COVID-19 and has not previously subscribed to the App for at least 14 days, such a user may voluntarily disclose their recent timestamped geolocation data, either via Google Timeline or Apple Location History data (if available), to be included in the daily dataset distribution.

Because of heightened risk of indoor disease transmission, indoor mapping technology in public and semi-public spaces must be enhanced with new spatial and fingerprint maps not already available to apps such as Google Maps or Apple Maps. Accurate and energy-efficient tracking in public and semi-public spaces, both indoor and outdoor, is achieved using a smartphone's native localization capabilities, which are based on GPS, WiFi fingerprinting, and Inertial Measurement Unit (IMU) sensors.

[1] While named "COVID-19 Geolocation App" for its initial release, the infrastructure created by this project is applicable to many infectious diseases, such as MERS or SARS.
[2] Period to be specified and regularly reviewed by the team's epidemiologist.

Implementation

Bluetooth proximity alone (see Singapore government's TraceTogether App, 2020, March), while useful for determining colocation with a confirmed COVID-19 case, does not record one's geolocation or environment and therefore cannot offer insight into potential surface-to-face transmission. WiFi fingerprinting enables indoor room and building level localization with high accuracy but requires a precise georeferenced survey of the unique WiFi signals that vary throughout a space. Architects, developers, and construction companies have detailed building floor plans that can be rapidly converted into open standards, such as the Indoor Mapping Data Format (IMDF). These floor plans are then used to generate the necessary surveys by walking through a space and recording WiFi signals. This process is guided by an app on a smartphone, and surveys can be scalably generated for buildings across entire cities through crowdsourcing of trained volunteers. In consultation with epidemiological experts, minor refinements to indoor mapping schema, such as adding a "touch/no touch" subclass on door and clear openings, are suggested to enhance the exposure risk assessment.

Implementation of this project is in three distinct phases. The COVID-19 Geolocation App, ISTM model, and indoor surveying are urgent and, if mobilized, could be built and deployed with sufficient accuracy in many locations globally in a matter of weeks. The second phase, over the next year or more, involves refining location services and indoor map data to account for specific epidemiological conditions. Advanced activity classification based on current Android- and iOS-supported accelerometer and gyro sensor measurements are ongoing topics of research in the computer science sensing community. When combined with indoor survey data and common existing geolocation technologies, sensed activities, such as pausing to open a door (i.e., touching a doorknob) or stepping onto a bus, could become a component of smartphone location services and passed to the COVID-19 Geolocation App for more effective exposure risk assessment. IMU activity classification at the population level with machine learning could then be used to refine and update WiFi fingerprints at specific key epidemiological locations, such as a door, and refine geolocation and proximity data on public transportation, including in bus and train compartments. Further, walkability databases exist or are under development in many cities and can either correct drift to sub-meter accuracy or allow further topological analysis. The third phase concerns the use of confirmed COVID-19 case geolocation data in longer-term epidemiological studies.

Different geographies have different urban densities (e.g., high-rise buildings, single-family detached homes), telecom markets (e.g., competition between providers), legal contexts (e.g., health data privacy laws), and technological capabilities (e.g., extensive WiFi fingerprinting, high-resolution Observed Time Difference Of Arrival/OTDOA capability). For simplicity and optimal execution, the risk assessment model should, whenever possible, use input only from a mobile device's vanilla location services (at its highest accuracy) rather than complicate the model with additional data. This is equally true for novel input location data, such as OTDOA, or data typically used afterwards to refine location accuracy, such as street vectors. It is better to be simple and overestimate user risk than to unnecessarily complicate the model. In limited contexts, such as Hong Kong (see Hong Kong Department of Health, 2020), where rigorous contact-tracing data has been made public, such data can easily be translated into comparable ISTM pointsets for increased localized effectiveness.

Privacy considerations

While the general population's privacy risk is minimal, there are potential privacy risks for those confirmed COVID-19 users who volunteer their location data. Before submitting their data, those users selectively redact home, work, and other locations within an area of their choice within the

App. In addition to encryption, confirmed COVID-19 users' submitted location data will be preprocessed server-side into an ISTM pointset that includes only those locations, timestamps, and durations at the lowest resolution necessary to assess exposure risk of colocation and surface adjacency in public and semi-public spaces. This location data must be stored, processed, and served by trusted entities. The duration that location services data (volunteered by confirmed COVID-19 users) is held on the server must be the shortest required duration, for example, a 14-day infectious period or 5-year epidemiological study. To increase both privacy and participation, a clear mechanism must be in place such that any user data submitted can be both accessed and destroyed promptly by the user at their request. No user should ever be required to disclose to any entity proof of App installation or the user's exposure risk calculated by the App.

Development team

The risk assessment model must synthesize both spatiotemporal epidemiological parameters and the varying accuracies of geolocation data in different contexts. This is a synthetic, iterative process that requires close collaboration and mutual agreement between experts in epidemiology, geostatistical analysis, and telecommunications technology. The minimum recommended development team must include:

- An epidemiologist who can advise on the rules for statistical optimization of travel paths and indoor data subclasses (e.g., IMU activity classification of most probable contact routes);
- A geostatistical analyst, who has extensive work with transport data, to code the cleaning and optimization of paths based on rules agreed by the entire team;
- A principled legal expert who has drafted or at least publicly argued for digital health data privacy restrictions; and
- A small experienced team of software engineers to code the App and manage the database. This team must have worked together on a globally distributed mainstream tracking App (for both iOS and Android systems), ideally one for health monitoring (e.g., running/outdoor sport). To ensure quality, it's important that the geostatistical analyst noted above be separate from the software/database team.

Data collection, processing, and use

For each voluntarily disclosed and confirmed COVID-19 case, a location services data history for 14 days is submitted to the server. This data is cleaned server-side to remove outliers and densified into an *infectious space-time map* (ISTM) pointset at an effective resolution. Both data cleaning and densification are dependent on the minimum risk assessment model requirements that effectively synthesize the current epidemiological understanding, location services data format and accuracy in use, and acceptable level of privacy of the COVID-19 cases. Seven variables are retained in the ISTM pointset: A unique ID, location (X-Y-Z), an interpolated timestamp, rate-of-travel (not direction), and any indoor map subclasses deemed useful for exposure risk assessment, such as a doorknob or eating surface. This is used for assessing user exposure risk (proximity and duration of co-location), which is computed locally on individual users' mobile devices to ensure no user-identifiable data (including location data) is uploaded to the server. A geographical subset of the ISTM pointset is downloaded by each user once per day to a local database on the user's mobile device, which is updated incrementally to reduce size for low-bandwidth areas. To protect privacy while keeping data transfer and processing efficient, this geographical subset could be a large (e.g., 50-kilometer) buffer around any fuzzy location. Further, the five variables at each ISTM point require relatively low precision (i.e., 1-meter X-Y-Z resolution, 1-minute timestamps). The rate-of-travel is an aggregate measure of all infected users at a given location and time duration, and, because infected cases are not in motion at all times, it is often near zero. Exposure risk is calculated under-the-hood (but publicly commented in the software documentation and code) based on the App's current risk assessment model. While user risk is calculated based on agreed rules, for clarity in communicating public health urgency, the App user interface should offer only the minimum information required to take a suggested action (e.g., "Yes, you are at risk based on your recent location history and are recommended to self-quarantine").

References

Hong Kong Department of Health. (2020, March). Data in Coronavirus Disease (COVID-19). Retrieved from https://data.gov.hk/en-data/dataset/hk-dh-chpsebcddr-novel-infectious-agent

World Health Organization (WHO). (2020, March 29). Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations.