REACT: Real-Time Contact Tracing and Risk Monitoring via Privacy-Enhanced Mobile Tracking

Yanan Da
Emory University
Atlanta, USA
yda3@emory.edu

Ritesh Ahuja
University of Southern California
Los Angeles, USA
riteshah@usc.edu

Li Xiong
Emory University
Atlanta, USA
lxiong@emory.edu

Cyrus Shahabi
University of Southern California
Los Angeles, USA
shahabi@usc.edu

***Abstract*—Contact tracing is an essential public health tool for controlling epidemic disease outbreaks such as the COVID-19 pandemic. Digital contact tracing using real-time locations or proximity of individuals can be used to significantly speed up and scale up contact tracing. In this demonstration, we present our system, REACT, for REAl-time Contact Tracing and risk monitoring via privacy-enhanced tracking of users' locations. With privacy enhancement that allows users to control and refine the precision with which their information will be collected and used, REACT will enable: 1) contact tracing of individuals who are exposed to infected cases and identification of hot-spot locations, 2) individual risk monitoring based on the locations they visit and their contact with others. In this paper, we demonstrate the procedure of contact tracing using our application and the utility of contact tracing given the protected locations.**

I. Introduction

More than 10 million people in the U.S. have been infected with the coronavirus (COVID-19) and more than 200,000 have died as of November 2020[1]. Nationwide, hundreds of thousands of new cases are still reported daily.

Contact tracing [1] is the quintessential public health tool for controlling epidemic disease outbreaks such as the COVID-19 pandemic. Tracing involves identifying all individuals who may have come into contact with an infected person and advising them to safely isolate or quarantine at home. Traditional contract tracing practices, such as those currently adopted by the CDC[2], require a contact tracer to assess an infected individual contacts by asking about his/her activities. This process, however, does not scale. It is time-consuming and ultimately infeasible for large scale contact tracing in a pandemic, as is the case for COVID-19. In addition, contact data collected in this way may be incomplete (limited to known contacts) or unreliable. Digital contact tracing using real-time locations or proximity of individuals can significantly speed-up and scale contact tracing, as demonstrated by many efforts in Asia and Europe [2]–[7].

User privacy is the critical issue in sharing real-time location traces of users for digital contact tracing. Uncontrolled sharing of users' whereabouts can lead to a wide range of attacks, from stalking and assault, to various privacy breaches that may disclose sensitive personal details such as one's health status, political or religious orientation, etc [8].

Many contact tracing applications use only Bluetooth-collected proximal pairings, not absolute geo-coordinates, to protect location privacy. Examples of this include official contact tracing apps from countries such as United Kingdom, Switzerland, and Germany. Of those apps, some keep the contact data locally in the user's phone while others upload the contact data to a central location (e.g., Singapore, Australia). However, ignoring absolute locations sacrifices the ability to estimate risk based on the type of the locations and identified hot spots and the ability to trace indirect contacts. A select few (e.g., Norway) collect both bluetooth contact data and GPS location data. This approach has led to privacy concerns and consequently a low adoption rate among citizens[3]. There have been apps that require mandatory location check-ins from citizens issued by governments like China [9]. While highly effective for containment interventions, these apps have also heightened concerns about surveillance and data abuse.

We believe a pandemic like COVID-19 requires a careful design of privacy protection—with public health benefits and privacy enhancement approaches—that optimizes the trade-offs (illustrated in Figure 1). Our contributions are summarized as follows.

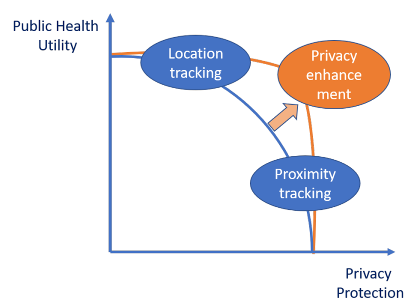


Fig. 1: Public Health Utility and Privacy Tradeoffs

- We present our application, REACT, for privacy-enhanced contact tracing using real-time locations.
- REACT protects users' location privacy by perturbing their exact geo-coordinates using Geo-Indistinguishability, a powerful location privacy definition

[1]https://covid.cdc.gov/covid-data-tracker
[2]https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/contact-tracing.html
[3]https://www.bbc.com/news/technology-52355028

that extends the popular protection model of differential privacy.

- Only perturbed locations are reported to the server. Inherently, there is utility loss associated to this process: contact tracing services will return information relevant to the reported location, instead of the actual one. However, the users can adjust their privacy levels (the precision of their uploaded locations) as their risk evolves.
- We propose an algorithm that models (as a probability) the possibility of an actual contact between an infected user and his/her contacts, given only their perturbed locations. We optimize to maximize this accuracy.
- We propose efficient implementations of the spatiotemporal queries for such contact tracing given the potentially large number of users.

II. Approach

A. Contact Tracing

Digital contact tracing using real-time locations can systematically identify all users who have been in contact with an infected case both directly and indirectly. There are three common types of transmission of COVID-19: 1) direct person-to-person transmission, i.e. in close contact with someone infected (simultaneous co-location); 2) fomite transmission, i.e. in contact with a contaminated surface or object at a location visited by someone infected earlier (lagged co-location) [10]; and 3) indirect person-to-person transmission by contact with someone who is earlier in contact with someone infected. Figure 2 illustrates four user trajectories where u_c has a confirmed infection at time t_4 and the three transmission scenarios: 1) u_1 via simultaneous co-location at t_2, 2) u_2 via lagged co-location at time t_4, and 3) u_3 via direct transmission from u_1 at t_3 (indirect transmission from u_c).

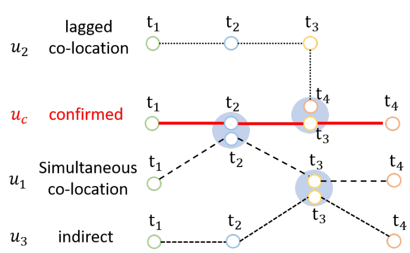


Fig. 2: Transmission Scenarios

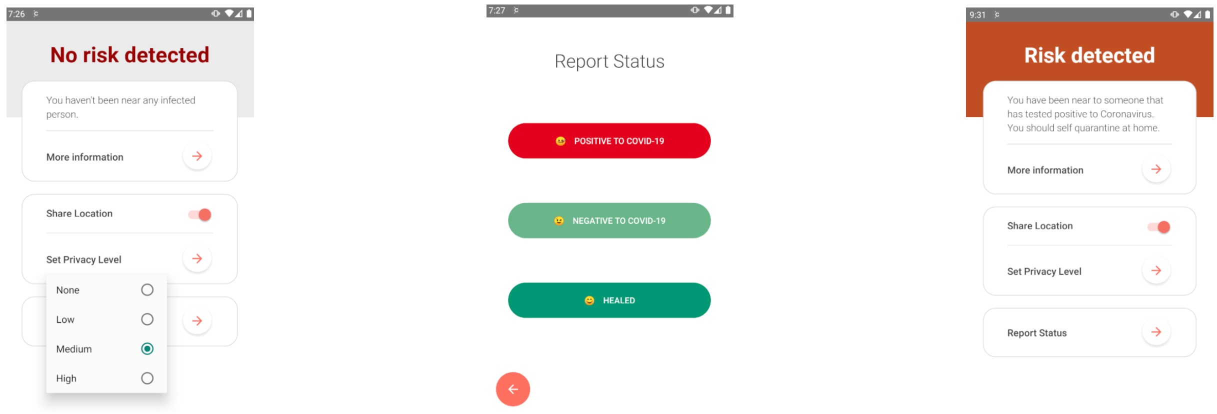
In REACT, we primarily consider the first two scenarios which are the primary ways of transmission. Assume the location trace of each user is represented as a sequence of *visits* $< u, s, t >$ (user u at location s at time t). Given a user u_c with confirmed infection, we find who have simultaneous or lagged co-location with u_c in a back tracing window (this can be parameterized, e.g. as 5 days which is the currently known median incubation period for COVID-19 [11]). Given a sequence of visits by a single user, this problem can be formulated as composition of *range queries* which given a location dataset, a spatial range and a time interval, returns all locations s that intersect with both the spatial range and the time interval. For each location s visited by u_c at time t in the time window, we can compute a range query using a spatial range of 6 feet centered at s and a time interval of up to 72 hours starting at t (the viable time of COVID-19 virus on surfaces [12]). These parameters can be adapted to characteristics of individual locations (e.g., interpersonal distances vary in closed/open spaces) and particular viruses (e.g., infectious spread range of aerosolized particle, as determined by research and CDC recommendations). Consider the example in Figure 2, the query given u_c will return u_1 and u_2. To scale such spatial queries, we utilize R-trees [13] to index the visits on the three dimensions: latitude, longitude and time.

B. Privacy Enhancements

To mitigate privacy risks and ensure widespread adoption, we give users full control over how their information will be collected. In particular, REACT allows users to control the precision and frequency with which their geo-coordinates are reported, controlled via an intuitive set of privacy setting options. User's geo-coordinates are perturbed using the geo-indistinguishability (GeoInd) privacy definition [14]–[17] based on users' choices of privacy level. Given a privacy parameter ε, and app users u_1 (and u_2, respectively), ε-GeoInd perturbation mechanism distorts their exact locations l_1 (l_2) to l'_1 (l'_2) by adding a spatial noise vector derived from a 2D Laplace distribution (with scale inversely proportional to ε). The challenge is then to accurately compute the range or reachability queries over the perturbed locations. We extend probabilistic techniques of [16] to calculate the range query over the pair of perturbed locations l'_1 and l'_2. Recall that the range query captures whether or not two users actually made a contact (parameterized as a reachable distance R), which is indicative of the risk of a potential transmission. The objective is then to calculate their reachability probability $p(d \le R|d')$, where d and d' are the Euclidean distances of their exact and perturbed location pairs, respectively.

The simplest way to approximate the reachability probability distribution $\Pr_{d,d'}$ (for a given R and ε) is by numeric computation using a publicly available locations dataset (since a closed form solution may not exist). First, a large number of user pairs are perturbed according to ε-GeoInd, and the resulting distances (actual distance d and the perturbed d') are recorded. Then a function $\mathcal{F}(d, d')$ maps the perturbed distances over discretized spatial ranges, e.g., in meters $[0...10), [10...20), ..., [500...\infty)$ to actual distances, such that the cumulative distribution of d for a given range of d' is approximated. This way the mapping function approximates the desired probability distribution $\mathcal{F}(d, d') \approx \Pr_{d,d'}(d \le R|d')$. With this approximate function, we can define a threshold $0 < \alpha < 1$, to consider a pair of users *in-contact* if the reachability probability between their perturbed locations is greater than α.

Given a reachable distance R and a threshold α, our contract tracing query works as follows: 1) determine the maximum d' that satisfies $\Pr_{d,d'}(d \le R|d') > \alpha$; 2) instead of R, the maximum d' (much more relaxed than R) is used to perform a



(a) Privacy Level Control (b) Report Status (c) Alert

Fig. 3: Mobile App Demonstration

range query on the indexed dataset; 3) the candidates returned by the range query are refined to get the final results. The first and second steps ensure that all the locations that are reachable with probability α are retrieved.

To further improve the precision of the contact tracing query, we can adopt a multi-stage privacy approach where users can adjust the privacy level of the location to be uploaded as their risk evolves. When the server identifies a user as a contact with a confirmed case, the user can choose to upload precise or less perturbed (lower privacy level) locations stored locally to confirm contact status.

III. Application Demonstration

The REACT[4] app extends an existing open source project named Covid Community Alert[5] with privacy features discussed in this work. REACT app collects proximal contacts (via Bluetooth) in addition to geo-coordinates (via GPS if permitted by user) for the purpose of contact tracing and assessing users' need to quarantine. It maintains the anonymity of its users by recording ephemeral device IDs that persist for the duration the app is installed and can be reset by user by re-installing the app. The overall architecture of the application is shown in Figure 4. We demonstrate the workflow of the client side and server side in the following sections.

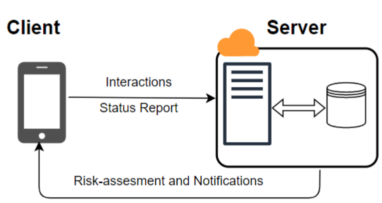


Fig. 4: Application Overview

A. Client Mobile App

We extended the mobile app with additional location privacy features. A UI element allows the user to select a desired privacy level as shown in Figure 3a for sharing his/her locations. This privacy level is interpreted as the level of perturbation that is applied to user's location before it is transmitted to the receiving server. We implemented the GeoInd based location perturbation with predefined privacy levels. If None is selected, the exact locations are reported to the server. The Share Location option allows the user to opt-in/out at any time. Another UI element provides the functionality for the users to self-report their COVID status as shown in Figure 3b.

The app works as follows. A user registers the device by sending a randomly generated device ID to the server on using the app for the first time (no personal information collected). The app scans for Bluetooth signals emanating from nearby devices and collects their IDs. The interaction information including devices IDs, timestamp, interaction duration, and the privacy-enhanced GPS location (if the user opts in) are sent to the back-end server. When a confirmed case is reported, the back-end server executes a *Contact Tracing Query* to find the potential contacts and update their at-risk score. If the risk score exceeds a preset level, an alert is relayed to these users as a notification on their device. As shown in Figure 3c, the main interface is also updated to reflect the notification status.

B. Contact Tracing Query

We will also demonstrate the contact tracing queries in the backend described in the previous section using a subset of the Gowalla Geo-social Network checkin dataset [18]. The dataset comprises a total of 3.6M check-ins within the US, from 54k unique users between Feb 2009 and Oct 2010. A randomly selected set of 500 users simulates the confirmed cases. Two check-ins are assumed to be co-located (representing a direct person-to-person transmission) if their Euclidean distance is within R and time interval is at most T. We set R to 25 meters and T to 20 minutes following previous work [19], [20]. We represent the privacy parameter ε as an easy-to-interpret average obfuscation value $\ell = \epsilon r$, where r is set to a constant 200 meters [14]. We vary privacy parameter ℓ from $ln2$ to $ln8$ (corresponding to average perturbation of 2.88 km to 0.96km, respectively).

We compare our approach against a straightforward application of GeoInd (dubbed the *oblivious* method) which counts two users as co-located when $d(l_1', l_2') \leq R$ where l_1', l_2'

[4]https://github.com/Emory-AIMS/react
[5]https://coronavirus-outbreak-control.github.io/web/

are perturbed locations. We evaluate two scenarios: the first, named U2U (uncertain to uncertain), assumes that both the infected user's location and other users' locations are perturbed; the second scenario, named U2E (uncertain to exact), assumes the infected user's locations is exact and other users' locations are perturbed. We compute precision (the fraction of correct co-location instances over all instances inferred), and recall (the fraction of correct colocations retrieved over the total count of co-locations in the original data) to evaluate the utility of the approaches.

Figure 5 (Figure 6) shows the results of the U2U (U2E, respectively) scenario where the privacy of the infected user's locations is set to $ln4$ and the privacy parameter of all other users varies from $ln2$ to $ln8$, corresponding to high to low privacy respectively. In both scenarios, our *probabilistic* method outperformed the *oblivious*. Both precision and recall increase with the decrease of user privacy level. This verifies that users with lower privacy requirement can receive more accurate risk estimation.

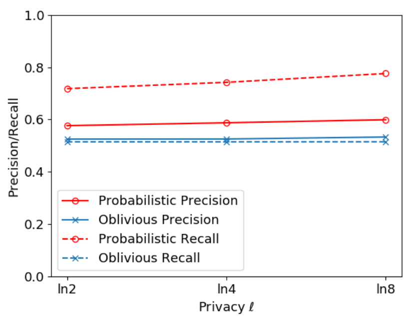


Fig. 5: U2U precision/recall by varying ℓ

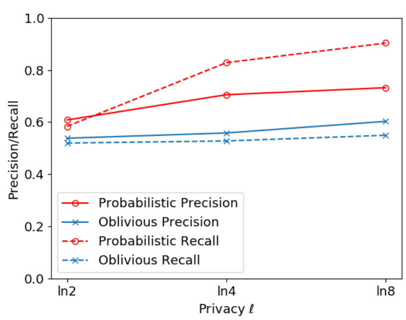


Fig. 6: U2E precision/recall by varying ℓ

Lastly, by exploiting spatio-temporal R-trees for computing co-location range queries, our algorithm is up to 300× more efficient, when compared to a nested-loop implementation that calculates the pairwise distance over all (infected user, other user) tuples.

For future work, we are investigating more fine-grained risk quantification approaches that take into account the type and risk factors of the locations the users visit in addition to the contact with infected users.

IV. Acknowledgement

This research has been funded in part by NSF grants CNS-2027783, CNS-1952192, IIS-1910950, CNS-2027794, NIH CTSA UL1TR002378, and unrestricted cash gifts from Microsoft. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsors.

References

[1] P. C. Erwin and R. C. Brownson, *Principles of Public Health Practice*. Cengage Learning, 2016.

[2] J. Jung, H. Jang, H. K. Kim, J. Kim, A. Kim, and K. P. Ko, "The importance of mandatory COVID-19 diagnostic testing prior to release from quarantine," *Journal of Korean medical science*, vol. 35, no. 34, p. e314, Aug. 2020.

[3] K. Nagarajan, M. Muniyandi, B. Palani, and S. Sellappan, "Social network analysis methods for exploring SARS-CoV-2 contact tracing data," *BMC medical research methodology*, vol. 20, no. 1, p. 233, 2020.

[4] C. M. Manauis, M. Loh, J. Kwan, J. Chua Mingzhou, H. J. Teo, D. Teng Kuan Peng, S. Vasoo Sushilan, Y. S. Leo, and A. Hou, "Bracing for impact: operational upshots from the national centre for infectious diseases screening centre (singapore) during the COVID-19 outbreak," *Journal of the American College of Emergency Physicians*, Jun. 2020.

[5] M. Jonker, E. De Bekker-Grob, J. Veldwijk, L. Goossens, S. Bour, and M. Rutten-Van Mölken, "COVID-19 contact-tracing apps: predicted uptake in the netherlands based on a discrete choice experiment," *JMIR mHealth and uHealth*, Aug. 2020.

[6] J. Peto, J. Carpenter, G. D. Smith, S. Duffy, R. Houlston, D. J. Hunter, K. McPherson, N. Pearce, P. Romer, P. Sasieni, and C. Turnbull, "Weekly COVID-19 testing with household quarantine and contact tracing is feasible and would probably end the epidemic," *Royal Society open science*, vol. 7, no. 6, p. 200915, Jun. 2020.

[7] M. Walrave, C. Waeterloos, and K. Ponnet, "Adoption of a contact tracing app for containing COVID-19: A health belief model approach," *JMIR public health and surveillance*, 2020.

[8] P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias, "Preventing location-based identity inference in anonymous spatial queries," *IEEE TKDE*, 2007.

[9] C. Thorbecke, "China launches coronavirus tracking app as death toll surpasses 1,000," ABC News, Feb. 2020, https://abcnews.go.com/Business/china-launches-app-combat-coronavirus-spread/story?id=68907706.

[10] A. N. Kraay, M. A. Hayashi, N. Hernandez-Ceron, I. H. Spicknall, M. C. Eisenberg, R. Meza, and J. N. Eisenberg, "Fomite-mediated transmission as a sufficient pathway: a comparative analysis across three viral pathogens," *BMC infectious diseases*, vol. 18, no. 1, p. 540, 2018.

[11] L. SA, G. KH, and e. a. Bi Q, "The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: Estimation and application," 2020.

[12] van Doremalen N, M. DH, and e. a. Holbrook MG, "Aerosol and surface stability of sars-cov-2 as compared with sars-cov-1," March 19, 2020.

[13] A. Guttman, "R-trees: A dynamic index structure for spatial searching," in *SIGMOD*, 1984, pp. 47–57.

[14] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *CCS*, 2013.

[15] Y. Xiao and L. Xiong, "Protecting locations with differential privacy under temporal correlations," in *CCS*, 2015, pp. 1298–1309.

[16] H. To, C. Shahabi, and L. Xiong, "Privacy-preserving online task assignment in spatial crowdsourcing with untrusted server," in *ICDE*, 2018.

[17] "Supporting both range queries and frequency estimation with local differential privacy," in *IEEE Communications and Network Security*.

[18] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *SIGKDD*, C. Apte, J. Ghosh, and P. Smyth, Eds., ACM. ACM, 2011, pp. 1082–1090.

[19] J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh, "Bridging the gap between physical location and online social networks," in *UbiComp*, 2010, pp. 119–128.

[20] H. Pham, C. Shahabi, and Y. Liu, "EBM: an entropy-based model to infer social strength from spatiotemporal data," in *SIGMOD*. ACM, 2013, pp. 265–276.