REACT: Real-time Contact Tracing and Risk Monitoring via Mobile Tracking (NSF RAPID Award)

Li Xiong, Computer Science, Emory University

Cyrus Shahabi, Computer Science, University of Southern California

EMORY UNIVERSITY

USC Viterbi
School of Engineering
Integrated Media Systems Center

UTHealth™
The University of Texas
Health Science Center at Houston

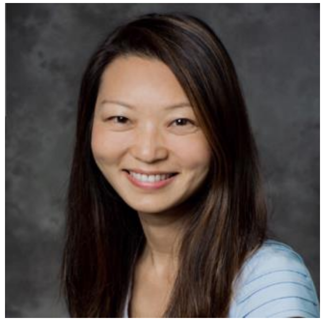
Project Team

Emory University

Li Xiong (CS and Biomedical Informatics)

Vicki Hertzberg (Nursing)

Lance Waller (Public Health)



University of Southern California

Cyrus Shahabi (CS, ECE, and Spatial Sciences)

University of Texas Health Science Center

Xiaoqian Jiang (Biomedical Informatics)

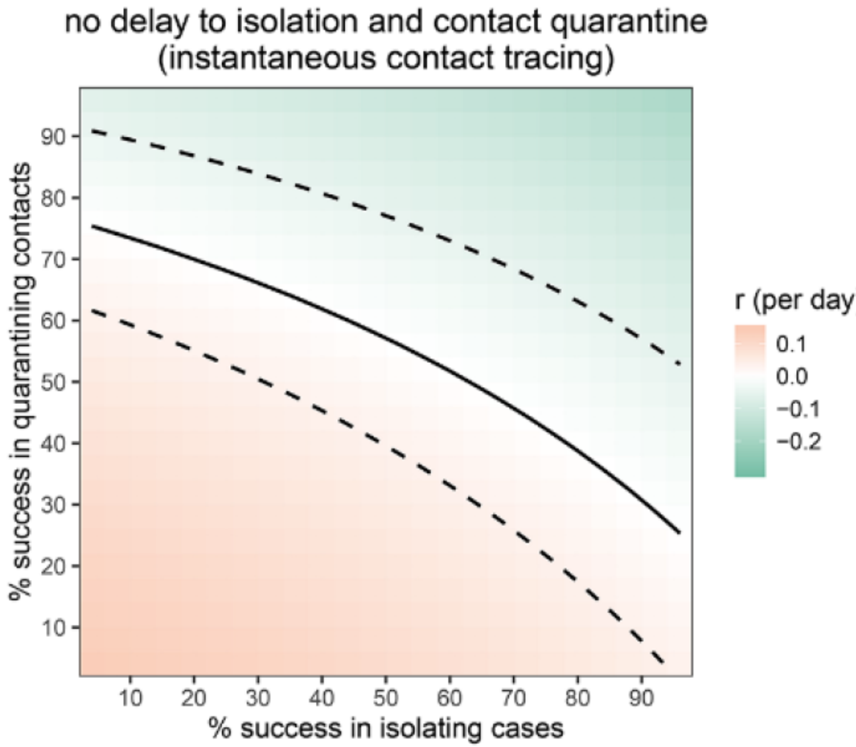
Amy Franklin (Biomedical Informatics)

Outline

- Project overview
- Privacy challenges and opportunities

Contact Tracing

- The incubation period is the time between infection and onset of symptoms
  - COVID-19: mean 5.5 days, median 5.2 days, standard deviation 2.1 days[1]
- The generation time is the time between the infection of the source and the infection of the recipient.
  - COVID-19: mean and median equal to 5.0 days, standard deviation of 1.9 days[2]
- For COVID-19, the incubation period is longer than generation time, so an infected person can transmit while pre-symptomatic
- Contact Tracing: interviewing a confirmed case (usually already symptomatic) to identify (and alert) his/her close contacts in the past x days (x=incubation time + symptomatic time; for COVID-19 14 days)
- Pluses: fomite transmission, indirect transmission
- Challenges: patient recall from memory, scale-up, latency
- Digital contact tracing: using mobile phones – promises to address all



Source: L. Ferretti et al., Science 10.1126/science.abb6936 (2020).

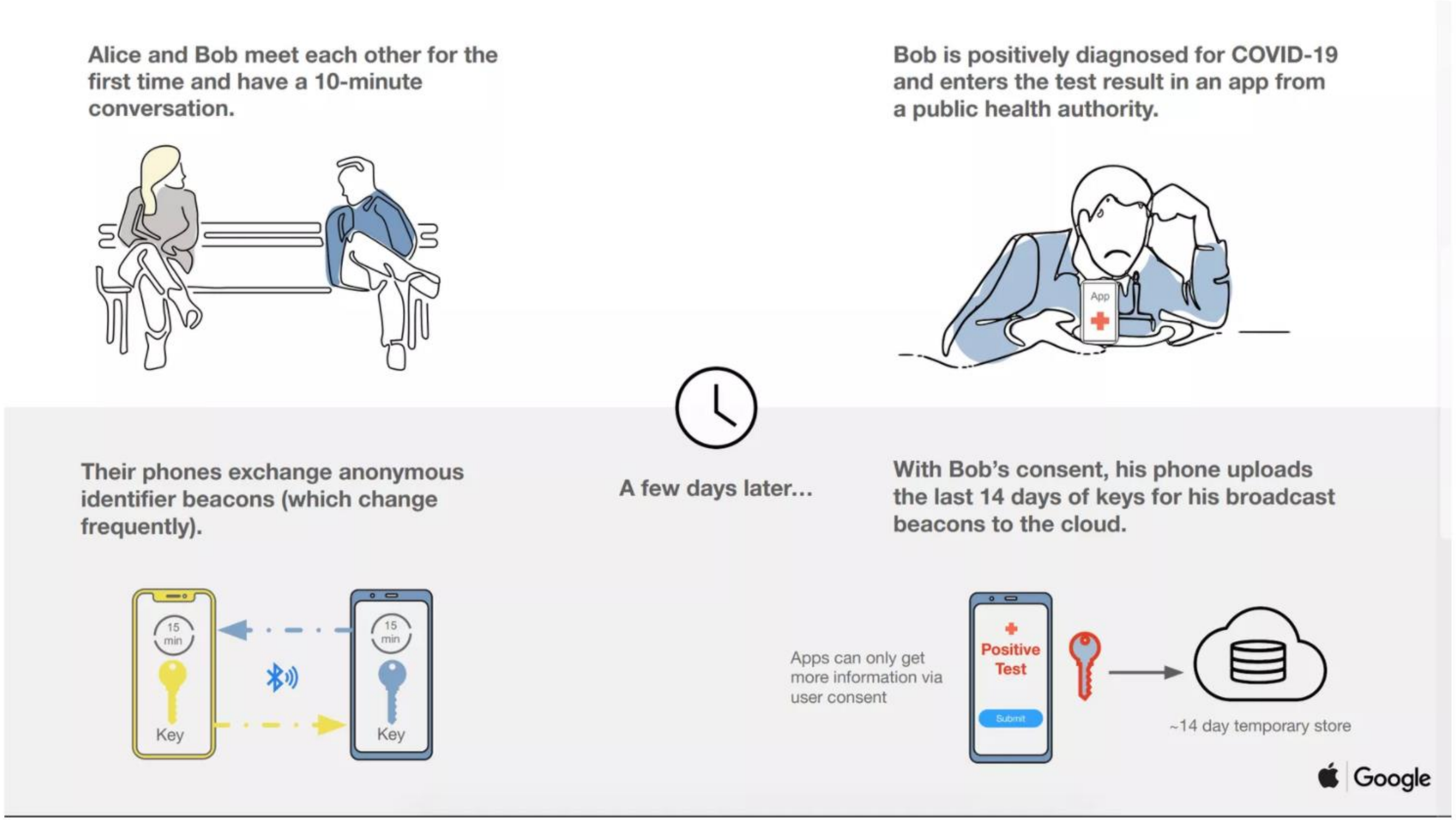
1. S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. Meredith, A. S. Azman, N. G. Reich, J. Lessler, The incubation period of 2019-nCoV from publicly reported confirmed cases: Estimation and application. medRxiv 2020.02.02.20020016 [preprint]. 4 February 2020.
2. L. Ferretti et al., Science 10.1126/science.abb6936 (2020).

Contact Tracing Apps (more than 50)

Contact-tracing Applications	Software Ready	Adoption ready	Privacy focus	Bluetooth	GPS	QR	iOS & Android	Open Source	medical certificate support	non profit	Centralized Logging	Server code	Adoption status	Web-link
Covid Community Alert	yes	yes	yes	yes	opt-in	no	yes	yes	yes	yes	yes	yes	Brazil (under consideration)	https://coronavirus-outbreak-control.github.io/web/index.html
OpenTrace (Singapore Govt.)	yes	support needed	partial	yes	no	no	yes	yes	no	yes	yes	yes	Singapore, Australia	https://github.com/opentrace-community
Coalition (Nodle)	yes	yes	yes	yes	no	no	yes	no	unclear	no (app free)	no	unclear	unknown	https://www.coalitionnetwork.org/
Private Kit: SafePaths (MIT)	alpha	no	yes	yes	yes	no	yes	yes	unclear	yes	yes	unclear	prototype	https://safepaths.mit.edu/
DP3T (EPFL and others)	alpha	no	yes	yes	no	no	yes	yes	yes	yes	no	yes	unknown	https://github.com/DP-3T/
CovidSafe (UW, Microsoft)	alpha	no	yes	yes	yes	no	Android only	yes	unclear	yes	yes	yes	prototype	https://covidsafe.cs.washington.edu/
Zerobase	alpha	no	yes	no	no	yes	yes	yes	no	yes	yes	yes	prototype	https://www.zerobase.io/

**Table courtesy of Prof. Bhaskar Krishnamachari*

Google/Apple proposed approach



Alice and Bob meet each other for the first time and have a 10-minute conversation.

Their phones exchange anonymous identifier beacons (which change frequently).

A few days later...

Bob is positively diagnosed for COVID-19 and enters the test result in an app from a public health authority.

With Bob's consent, his phone uploads the last 14 days of keys for his broadcast beacons to the cloud.

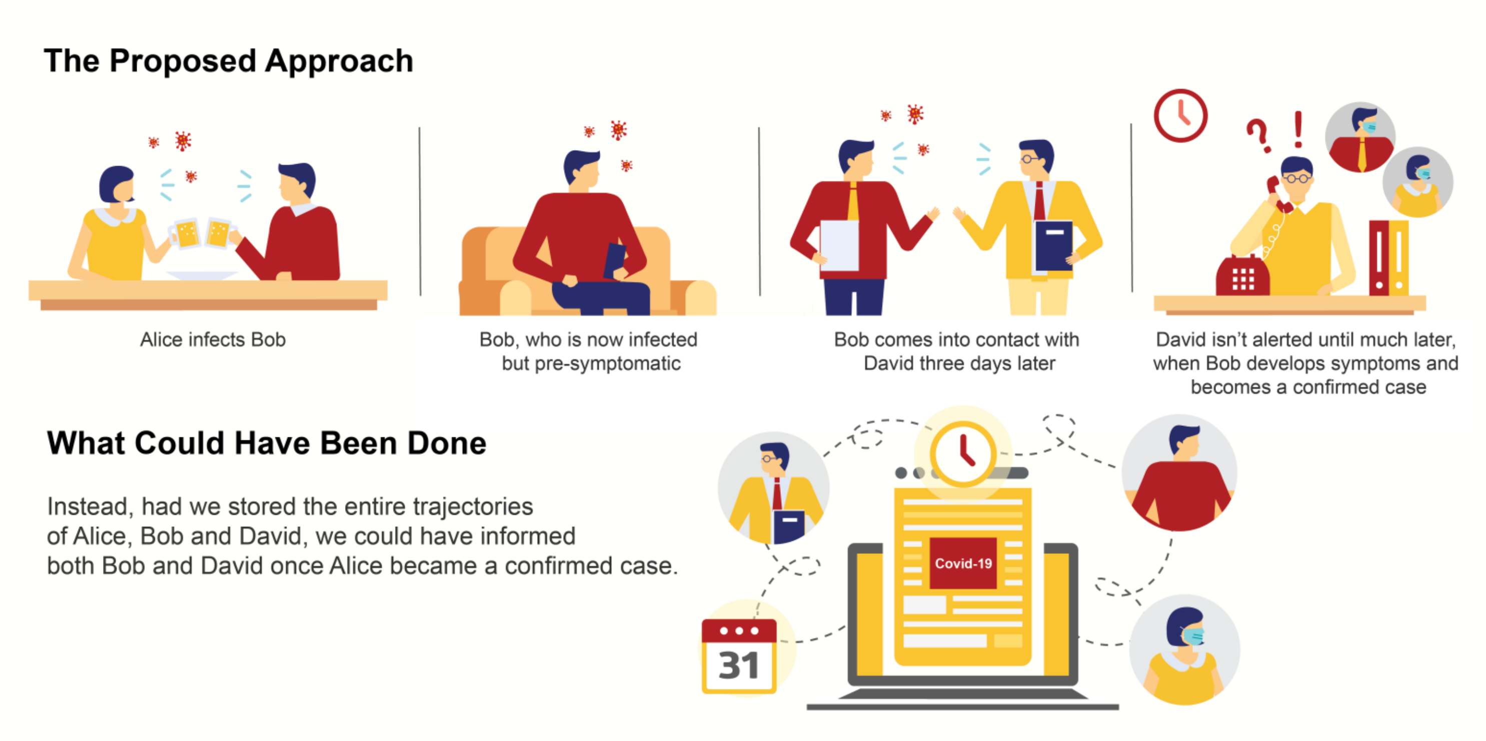
Apps can only get more information via user consent

~14 day temporary store

Apple/Google

Limitations of the approach

- Can’t detect fomites transmission
- Can’t detect indirect trans.
- Potentially high false positives
- By design is alert based
  - Too many false alarms; turning numb to them
- No centralized data collection
  - To inform policy, spread models, etc.



The Proposed Approach

Alice infects Bob

Bob, who is now infected but pre-symptomatic

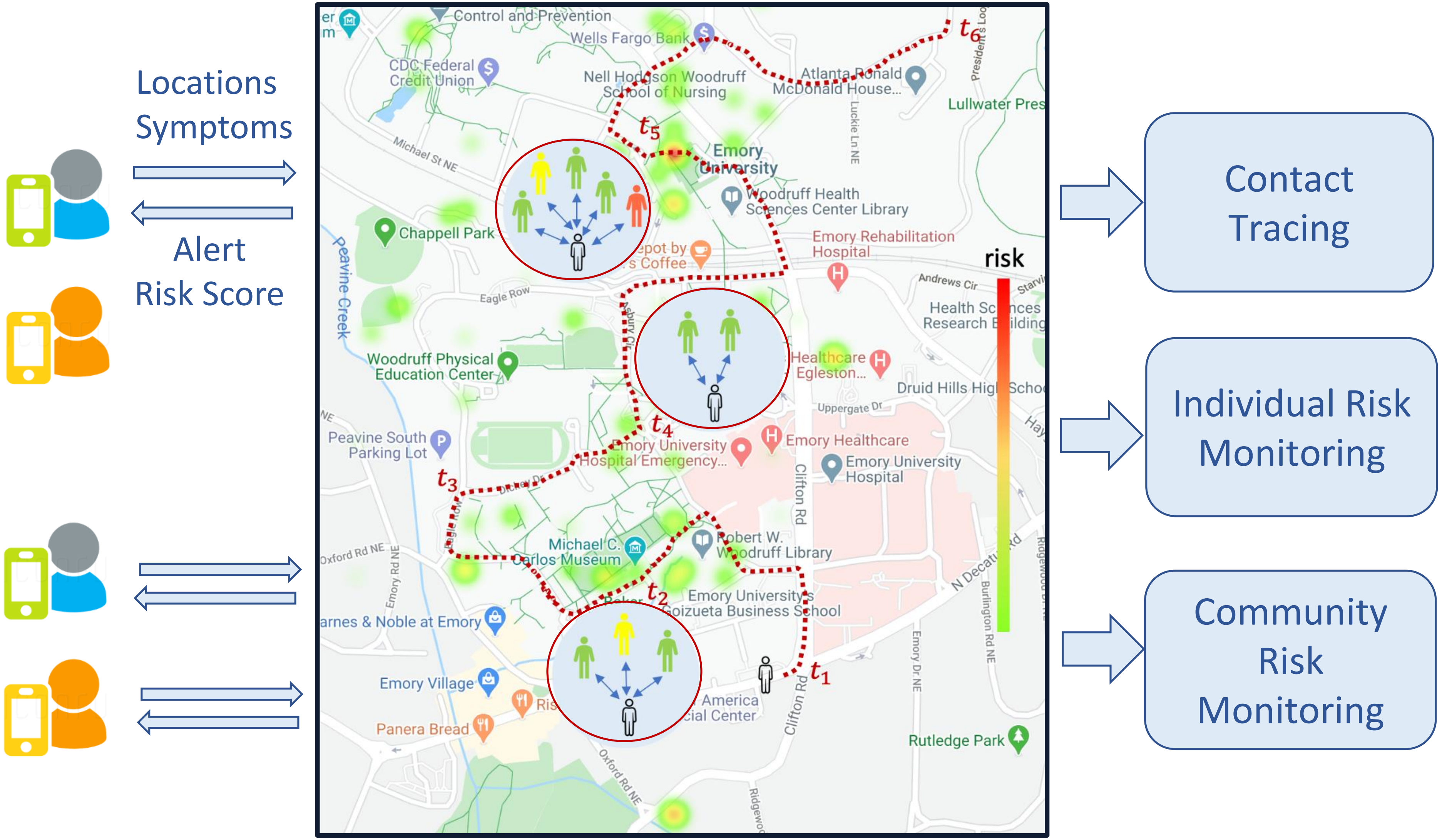
Bob comes into contact with David three days later

David isn’t alerted until much later, when Bob develops symptoms and becomes a confirmed case

What Could Have Been Done

Instead, had we stored the entire trajectories of Alice, Bob and David, we could have informed both Bob and David once Alice became a confirmed case.

REACT: Real-time Contact Tracing and Risk Monitoring via Mobile Tracking



Locations
Symptoms

Alert
Risk Score

Contact Tracing

Individual Risk Monitoring

Community Risk Monitoring

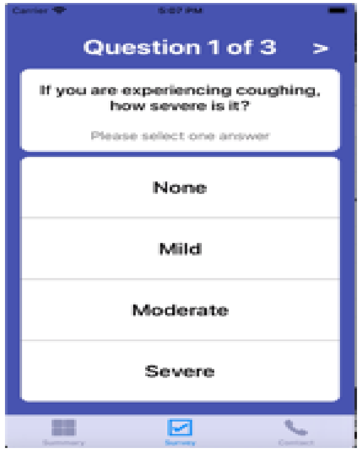
t_1
t_2
t_3
t_4
t_5
t_6
risk
Control and Prevention
Wells Fargo Bank
CDC Federal Credit Union
Nell Hodgson Woodruff School of Nursing
Atlanta Ronald McDonald House...
Lullwater Pres
Emory University
Woodruff Health Sciences Center Library
Chappell Park
Emory Rehabilitation Hospital
Health Sciences Research Building
Peavine Creek
Woodruff Physical Education Center
Healthcare Egleston...
Druid Hills High School
Peavine South Parking Lot
Emory University Hospital Emergency...
Emory Healthcare
Emory University Hospital
Clifton Rd
Michael C. Carlos Museum
Robert W. Woodruff Library
Emory University's Goizueta Business School
Barnes & Noble at Emory
Emory Village
Panera Bread
Rutledge Park
Oxford Rd NE
N Decatur Rd
Eagle Row
Michael St NE

Contact tracing and beyond

- Community app
  - Opt-in and voluntarily
  - Collect more data (e.g., locations, traces, durations), plus contact
- Contact tracing based on more data
  - Less false positives, detect fomites and indirect transmissions
- Risk Monitoring: In-app risk score [0,1]
  - Learning based approach
  - No more alerts, more consistent with how we use apps
- Community risk monitoring
  - Social sensors-based approach (friends group vs random group)
- Privacy enhancements
- Other usage of the collected data
  - Location risk scores
  - Inform spread models

Spatiotemporal Analysis

- App to collect:
  - Symptoms, Locations, Traces, Durations, Histories
- Spatiotemporal queries enable detection of
  - Fomites transmission
  - Indirect transmission



Question 1 of 3 >

If you are experiencing coughing, how severe is it?

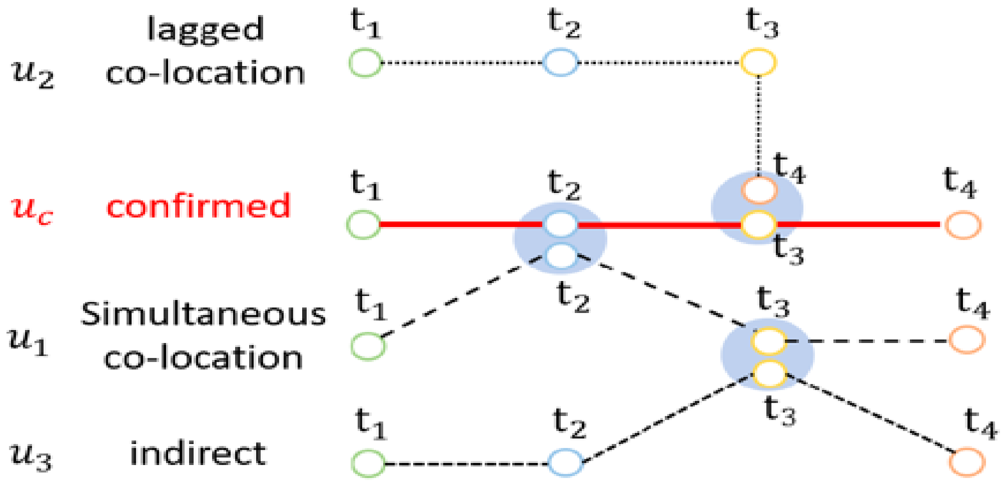
Please select one answer

None

Mild

Moderate

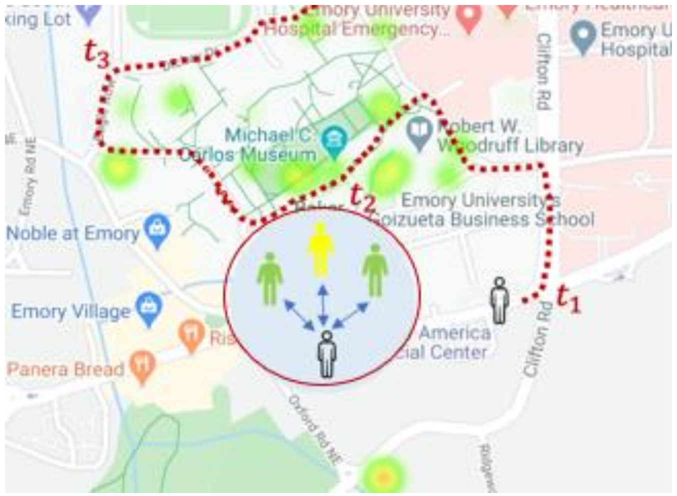
Severe



H. Shirani-Mehr, F. B. Kashani, and C. Shahabi. Efficient reachability query evaluation in large spatiotemporal contact datasets. PVLDB, 5(9):848–859, 2012.

Risk Analysis

- Individual risk score:
    - Risks of contacted individuals
    - Risks of visited locations
    - Durations of contacts
    - Location of contacts
    - Social strength of contacts
    - Auxiliary data (if available): EMR, demographics



- Use social relationship inferred from people's historical trajectories

H. Pham, C. Shahabi, and Y. Liu. EBM: an entropy-based model to infer social strength from spatiotemporal data. In Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2013, New York, NY, ACM, 2013.

Outline

- Project overview
- Privacy challenges and opportunities

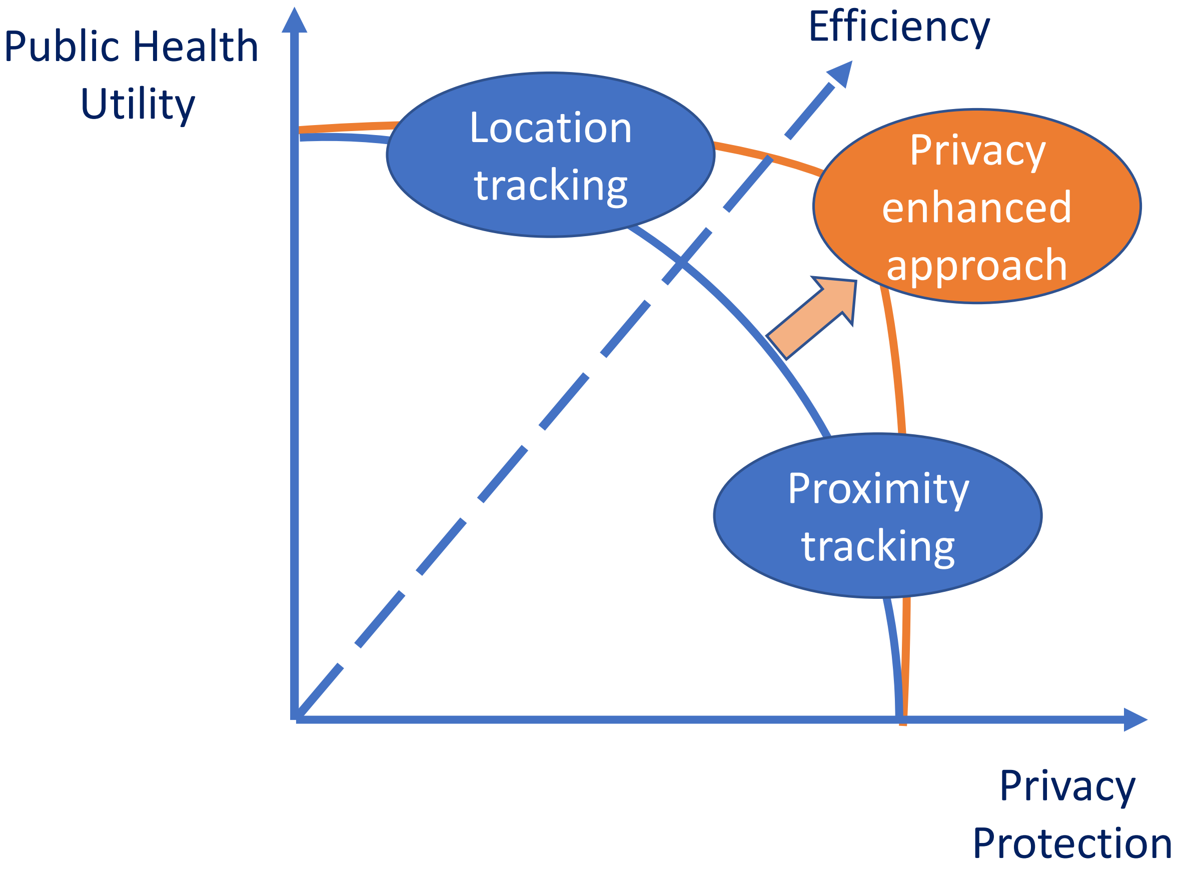
Security/Privacy Considerations

- Consent – informed, voluntary, and opt-in consent
- Transparency - what data is collected, how it will be used and by whom
- Minimization - collect the least amount of information needed
- Information security – audits/penetration testing, false reports/denial of service, the system should be secure (vetted by security experts)
- Addressing bias – not leave out marginalized groups and introduce false positives for certain users
- Expiration - not meant to last beyond the COVID19 outbreak

https://www.brookings.edu/techstream/inaccurate-and-insecure-why-contact-tracing-apps-could-be-a-disaster/
https://www.eff.org/deeplinks/2020/04/challenge-proximity-apps-covid-19-contact-tracing

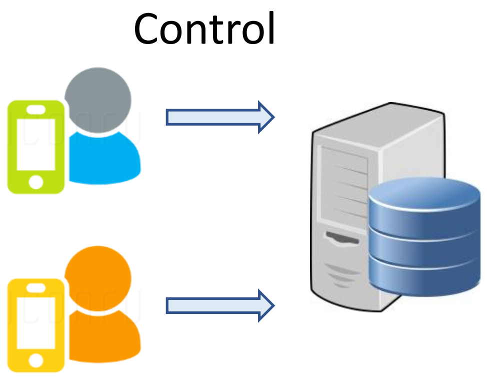
Privacy Challenges

- How to balance and enhance the privacy and utility (and efficiency) tradeoff?
- How to quantify privacy risks/protection and utility?



Basic Privacy Enhancements

- Location traces stored locally
- Users have full control of data collected by the server
  - Frequency (or manual check-in)
  - Precision of locations
  - Subset of locations (e.g. public/dense locations)



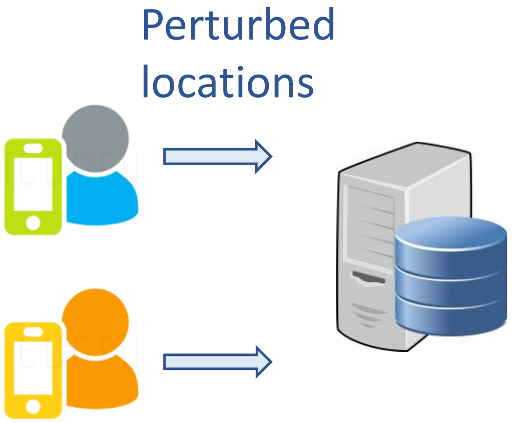
- Two stage (iterative) privacy approach
  - Server identify possible contacts/compute individual risks based on collected data
  - Alerted (or high risk) users can choose to upload full/precise locations for refinement
- Research opportunities
  - How to formally measure the privacy risks for users
  - How to balance false positive/false negatives given imprecise/incomplete data
  - How to incentivize users to contribute more data (e.g. by showing the value of their data for global contact tracing or risk analysis)

Potential Privacy Enhancement Mechanisms

- Location perturbation (geo-indistinguishability and variants)
- Searchable encryption
- Federated learning
- Machine learning with differential privacy
- Data sharing with differential privacy

Location perturbation

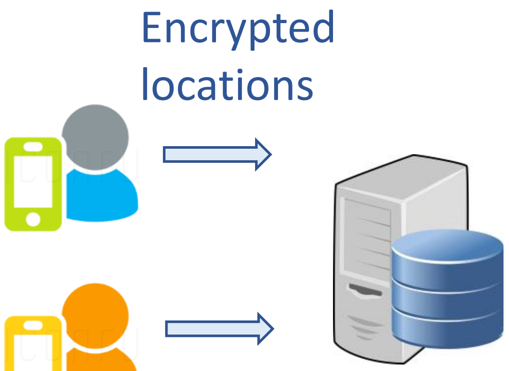
- Use: contract tracing and risk monitoring using perturbed locations
- Challenges: inaccuracy
- Research opportunities:
  - Interactive/iterative approach - perturbed locations for initial computation and precise locations for refinement; value-driven data collection
  - Input sensitivity and context - less perturbation for not-so-sensitive but more important locations for risk monitoring (e.g. grocery stores)
  - Utility metrics - contact tracing (individual queries) vs. other uses (e.g. aggregate location visits for global risk models)
  - Privacy quantification - additional adversarial knowledge (e.g. mobility patterns); spatiotemporal pattern protection (e.g. commute patterns) vs. location protection



Hien To, Cyrus Shahabi, Li Xiong. Privacy-Preserving Online Task Assignment in Spatial Crowdsourcing with Untrusted Server. 34th IEEE International Conference on Data Engineering (ICDE), 2018
Xiaolan Gu, Ming Li, Li Xiong and Yang Cao. Providing Input-Discriminative Protection for Local Differential Privacy. IEEE International Conference on Data Engineering (ICDE), 2020
Yang Cao, Yonghui Xiao, Li Xiong, Liquan Bai, Masatoshi Yoshikawa. Protecting Spatiotemporal Event Privacy in Continuous Location-Based Services. IEEE Transactions on Data and Knowledge Engineering (TKDE), 2020
Xiaolan Gu, Ming Li, Yang Cao and Li Xiong, Privacy-Preserving Range Queries and Frequency Estimation with Geo-indistinguishability. The 7th IEEE Conference on Communications and Network Security (CNS), 2019
R. Ahuja, G. Ghinita, and C. Shahabi. A utility-preserving and scalable technique for protecting location data with geo-indistinguishability. EDBT 2019

Searchable encryption

- Use: contact tracing using encrypted locations (centralized or decentralized)
- Challenges:
  - High computation cost
  - Difficult to support fomite transmission and indirect transmission; and incorporate other information such as location semantics
- Opportunities:
  - Optimize computation efficiency w/ specially designed spatial indexes
  - Specialized protocols for more complex contact tracing and risk modeling
  - Hybrid approach that combine location perturbation with searchable encryption – e.g. use imprecise/perturbed locations for precomputation and use encrypted computation for refinement

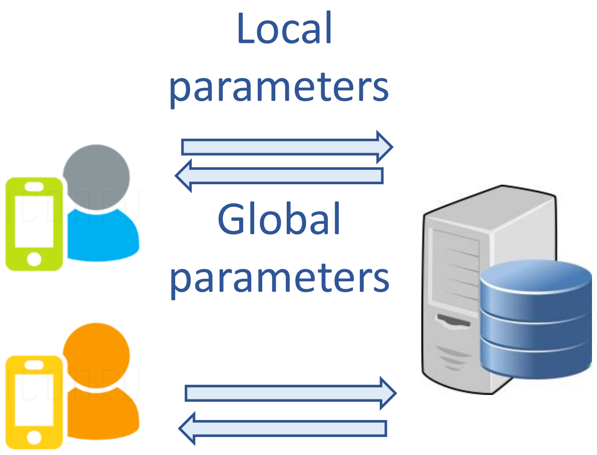


S. Shaham, G. Ghinita, and C. Shahabi. Enhancing the Performance of Spatial Queries on Encrypted Data through Graph Embedding. To appear in the 34th Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'20) June, 2020

Boyang Wang, Ming Li, Li Xiong. FastGeo: Efficient Geometric Range Queries on Encrypted Spatial Data. IEEE Transactions on Dependable and Secure Computing (TDSC), 16(2), 2019

Federated Learning

- Use: learn global risk models while maintaining data on local devices and sharing only model parameters
- Challenges
  - Cross-device learning may not be feasible - how to train local risk model with limited individual data
- Opportunities
  - Cross-institutional learning that incorporates both global factors (demographic, disease specific) and local (regional) factors

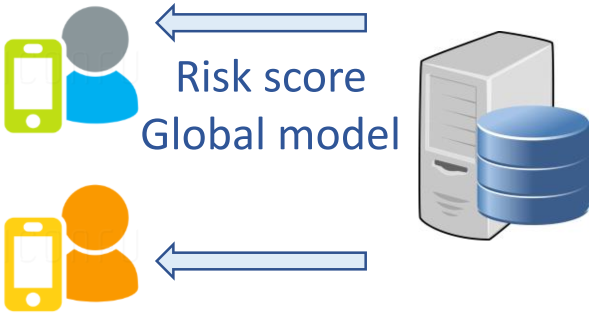


Local parameters

Global parameters

Machine Learning with Differential Privacy

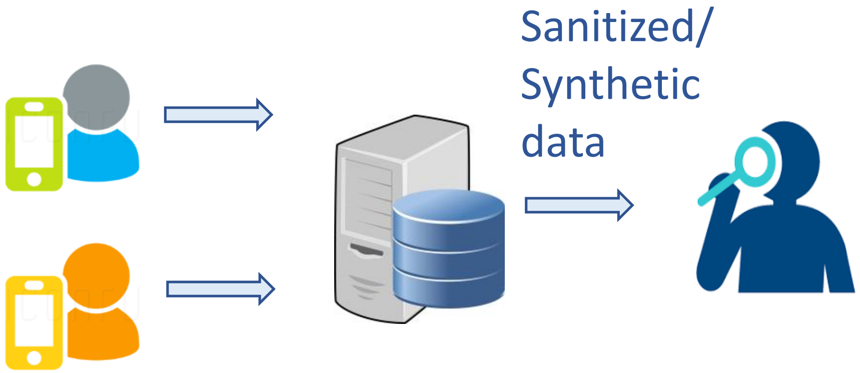
- Use: ensure risk scores and global models do not disclose other users' information
- Challenges
  - Inaccuracy/noise injected into the model
- Opportunities
  - Enhance privacy/accuracy tradeoff, particularly temporal models specific to real time risk monitoring
  - Practical impact on protections against attacks



R. Ahuja, G. Ghinita, and C. Shahabi. Differentially-private next location prediction with neural networks. EDBT 2020, Copenhagen, Denmark, March 28-Apr 2, 2020

Data sharing with Differential Privacy

- Use: sharing collected location (co-location) and symptoms data with researchers
- Challenges
  - How to preserve the spatiotemporal patterns/dependencies that are key for disease spread studies
- Opportunities
  - Mechanisms enhancing utility for spatiotemporal data
  - Quantification of privacy risks due to spatiotemporal correlations



He, Xi & Cormode, Graham & Machanavajjhala, Ashwin & Procopiuc, Cecilia & Srivastava, Divesh. DPT: Differentially Private Trajectory Synthesis Using Hierarchical Reference Systems. VLDB 2015

Kun Ouyang, Reza Shokri, David S. Rosenblum, Wenzhuo Yang. A Non-Parametric Generative Model for Human Trajectories. IJCAI 2018

Shengzhi Xu, Sen Su, Xiang Cheng, Zhengyi Li, Li Xiong. Differentially Private Frequent Sequence Mining. IEEE Transactions on Data and Knowledge Engineering (TKDE), 2016

Our Relevant Privacy Research

- NSF: Rigorous and Customizable Spatiotemporal Privacy for Location Based Applications (Xiong)
- NSF: PE4GQ - Practical Encryption for Geospatial Queries on Private Data (Shahabi)
- NIH: Decentralized differentially-private methods for dynamic data release and analysis (Jiang and Xiong)

Thank you

REACT: Real-time Contact Tracing and Risk Monitoring via Mobile Tracking

EMORY UNIVERSITY

USC Viterbi School of Engineering
Integrated Media Systems Center

UTHealth™
The University of Texas
Health Science Center at Houston