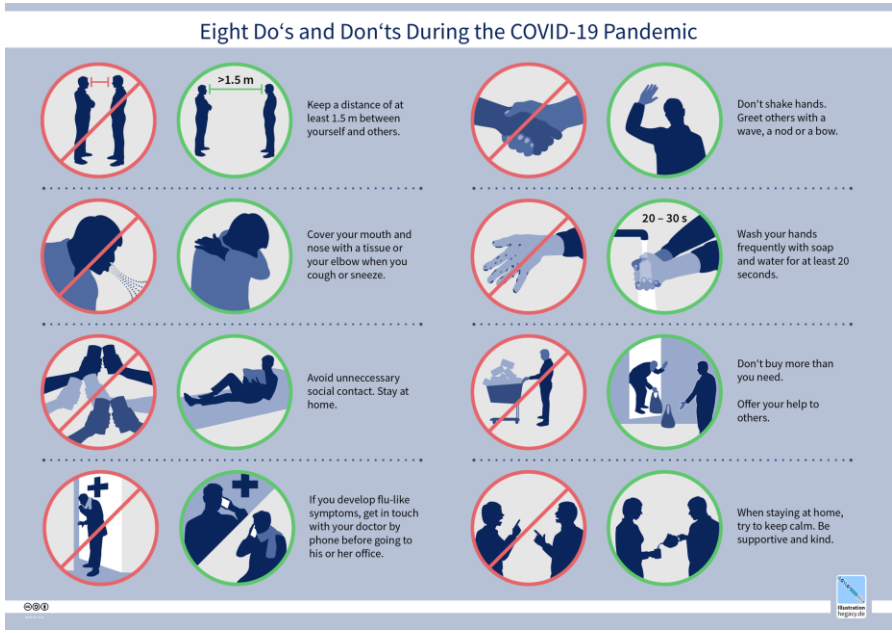
Anonymous Collocation Discovery: Harnessing Privacy to Tame the Coronavirus

Ran Canetti, Ari Trachtenberg, Mayank Varia

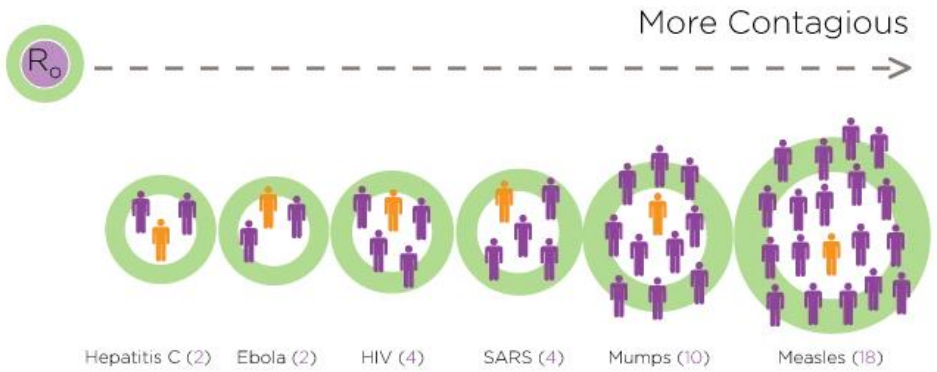
https://arxiv.org/abs/2003.13670



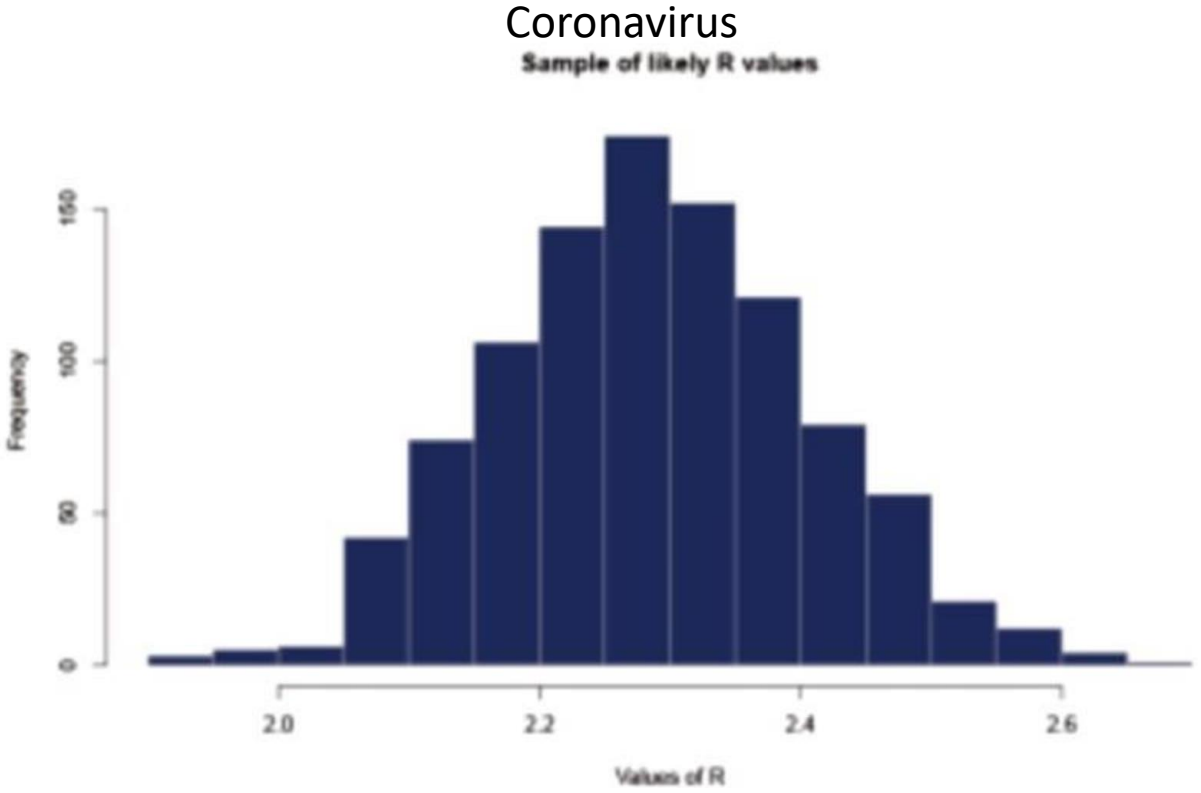
https://commons.wikimedia.org/wiki/File:14_Hegasy_COVID-19_Q_EN.png

The problem: R_0

= # of people who will catch the disease from one contagious person



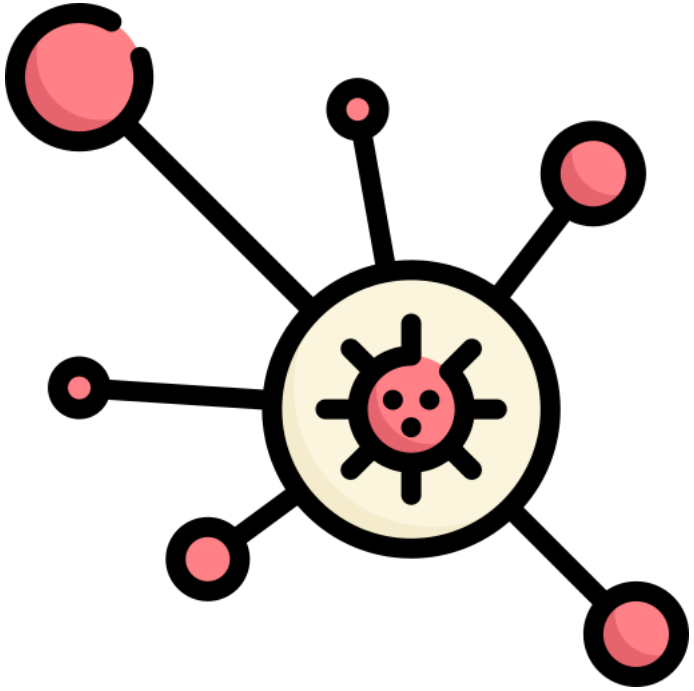
healthline.com



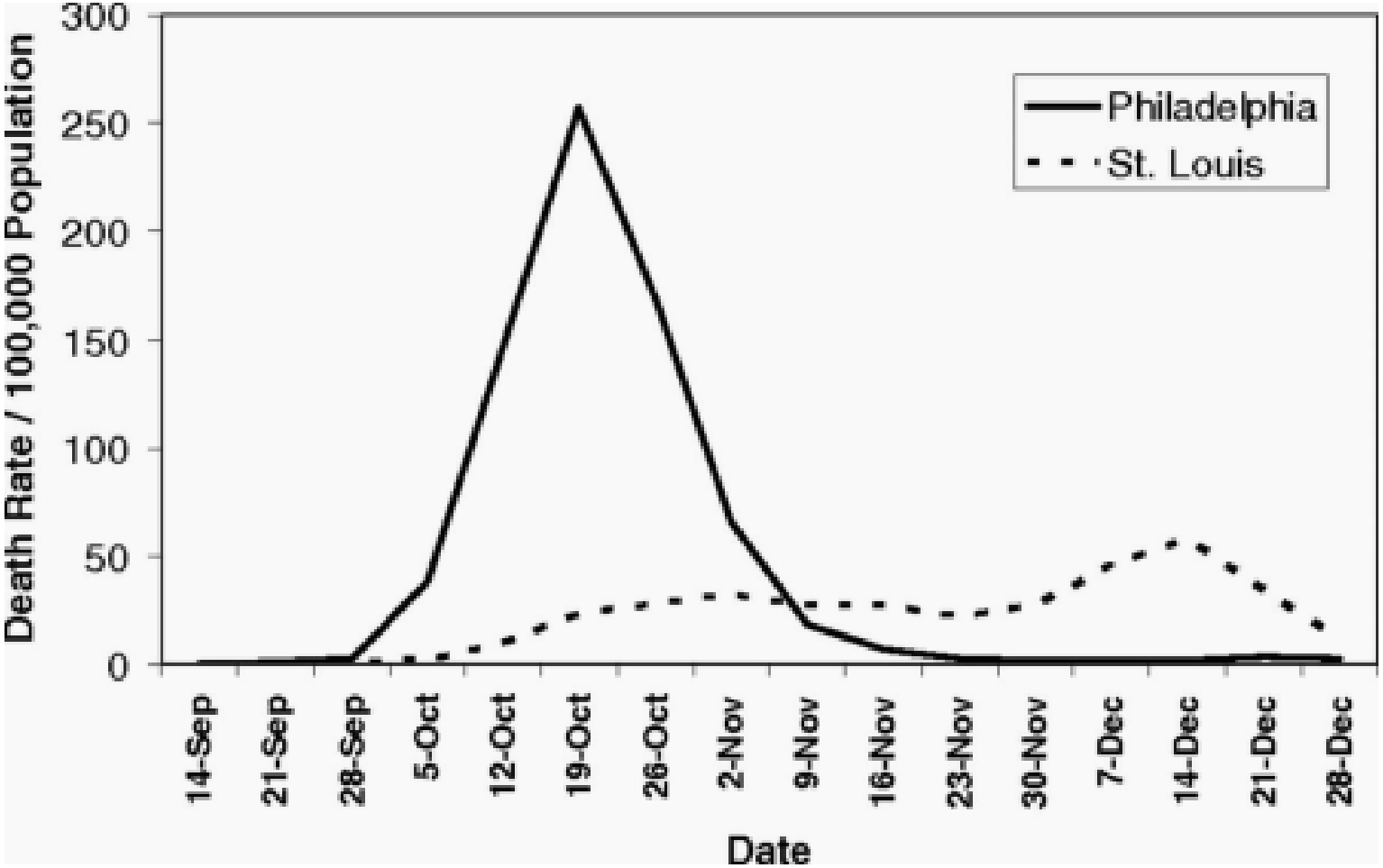
Sheng Zhang, et al. Estimation of the reproductive number of novel coronavirus (covid-19) and the probable outbreak size on the diamond princess cruise ship: A data-driven analysis. International Journal of Infectious Diseases, 2020.

Mitigation

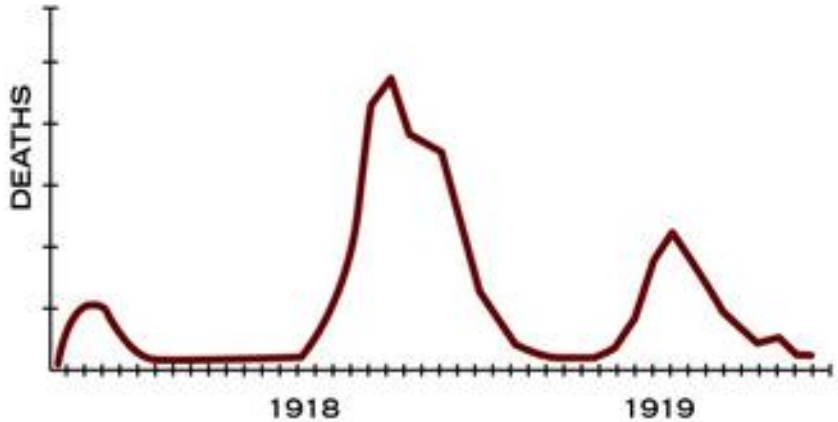
- **Spread continues until:**
  - *Starvation*
    - too hard to find victims
      - coupon-collector problem
    - herd immunity
    - ~50% of the population for $R_0 \sim 2$
  - *Vaccination*
    - 12-18 months away



Control - 1918 Pandemic



https://qz.com/1816060/a-chart-of-the-1918-spanish-flu-shows-why-social-distancing-works/



https://www.cdc.gov/flu/pandemic-resources/1918-commemoration/three-waves.htm

Minimizing infections:

- **General Quarantine**
  - Easier to implement
  - Requires complete coopeartion
  - Affects economy, psychology
  - Eventually loses effect
- **Targeted Quarantine**
  - Extensive testing
  - Timely alert and isolation of infections
  - COVID:
    - People are contagious while asymptomatic

College students relax and have fun during their Spring Break. (AP Photo/Alan Diaz)

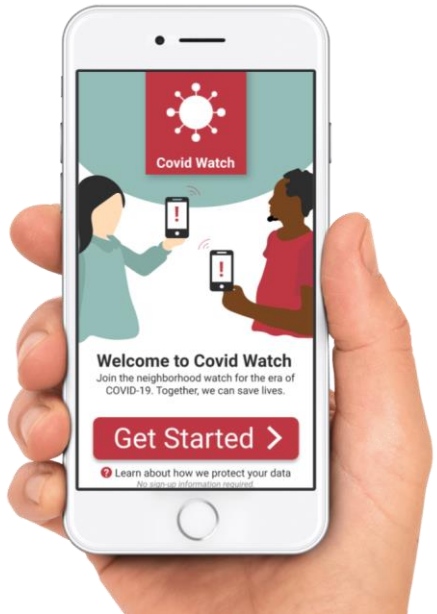
System goals:

- **Open participation**
  - Voluntary
  - Enter and leave at will
- **Simplicity**
  - Easy to understand
    - may affect adoption!
  - Easy to implement
  - Easy to verify
- **Decentralization**
  - No central personal information
  - Cannot aggregate databases
- **Low infrastructure**
  - Deployment must be fast

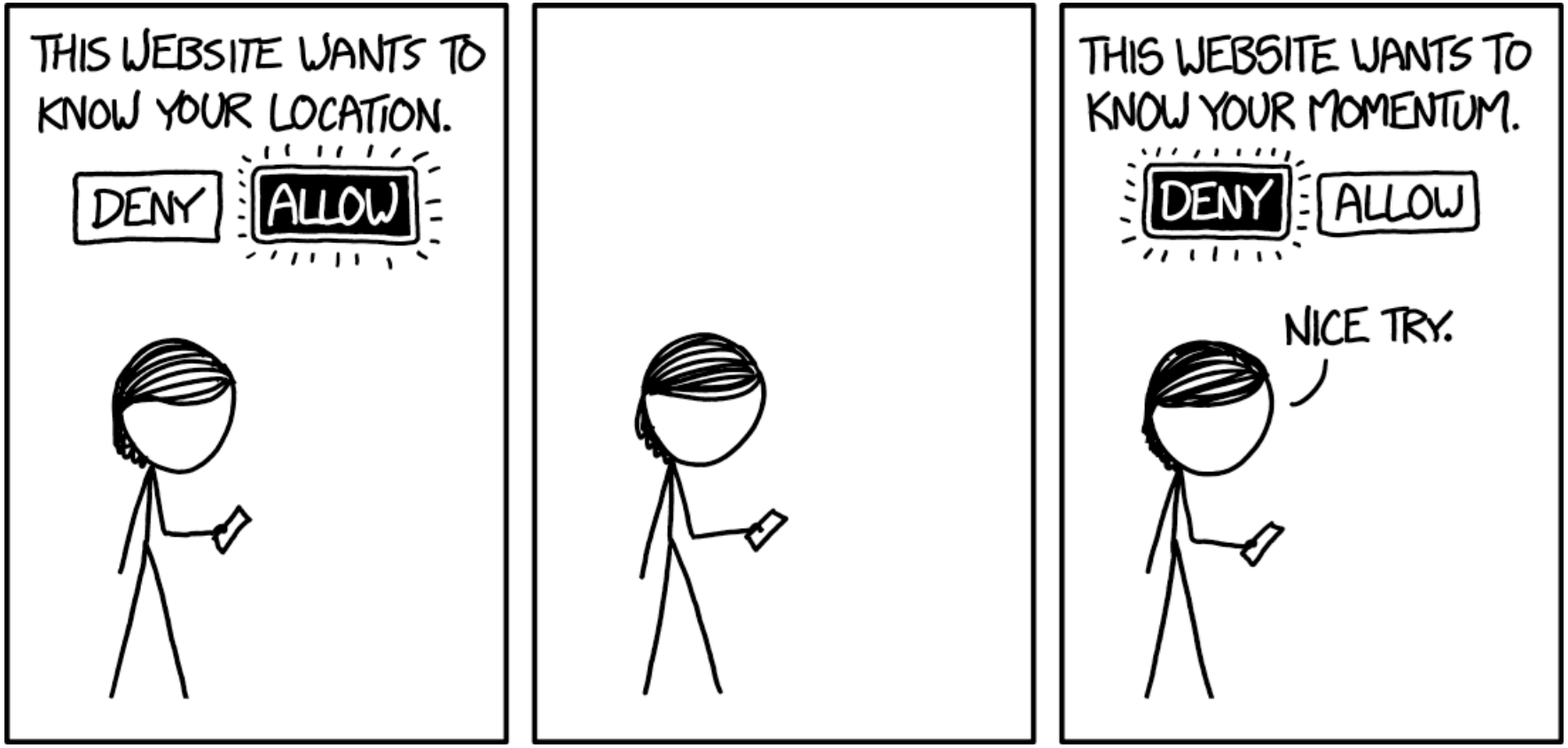
Existing systems:

- **China, Taiwan, and South Korea:**
  - Central aggregation of cellphone data
  - No public details
- **Singapore**
  - Bluetooth contacts + GPS location history
  - Privacy from other users
  - No privacy from government
- **Israel**
  - Privacy until infected
  - Full location history of infected party is shared
- **Covid-watch, MIT**
  - Simialr to this scheme
- **Lindell and Green**
  - Brighttalk on scientific and political challenges

TraceTogether



Location history – why do we care?



https://xkcd.com/1473/

Location Leakage … *Why should we care?*

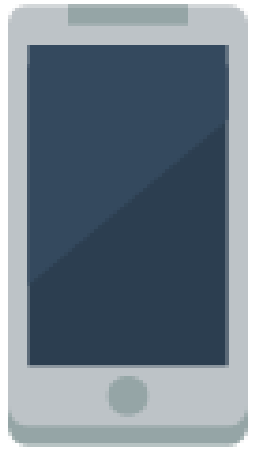
- Work times
- Friends
- Medical issues
- Political/religious interests
- What you buy / consider

COVID status … *Why should we care?*

- Social shaming
- Employment risk
- Insurance
- Social score

High level idea:

- **Joe user**
  - Broadcasts random tokens.
    - Short-range - Bluetooth
    - Approximates infection risk distance
    - Rotated at regular intervals
  - Listens for other broadcasts.
    - Checks received tokens against infected registry
- **Potentially sick user**
  - Gets tested.
  - If positive, uploads broadcasts to central registry

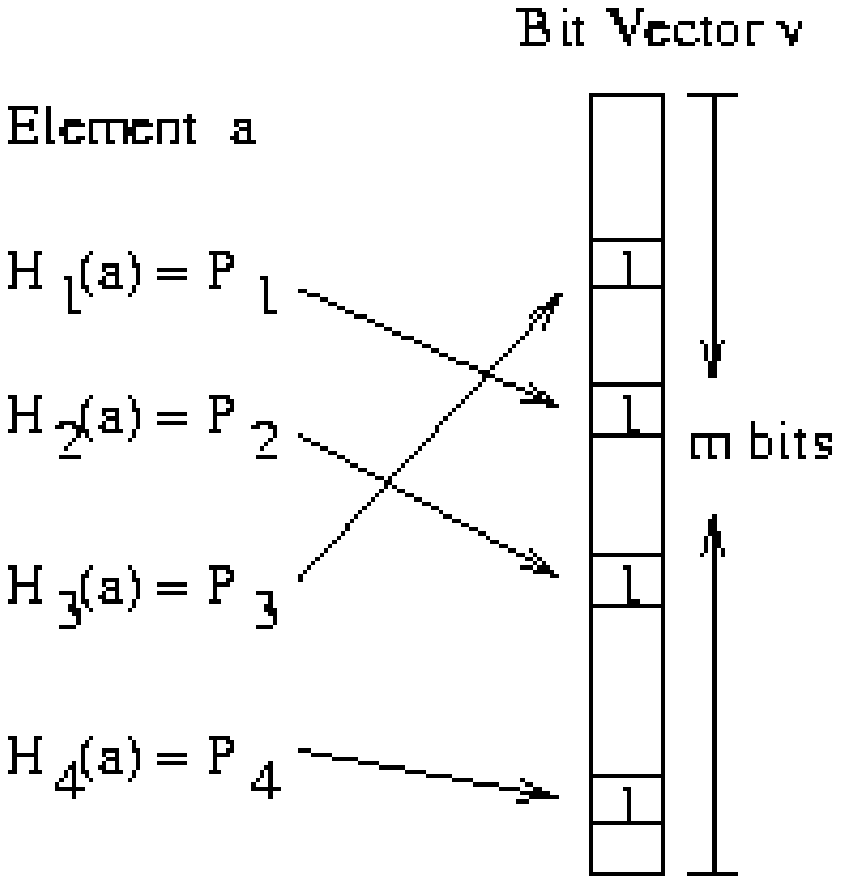


Parameters:

- **Time epoch (a tick)**
  - Long enough to be reproduced reliably
  - Short enough to produce privacy
- **Retention time**
  - How long data is retained (14 days?)
- **Update interval**
  - How often to contact the registry

Implementation using Bloom filters

- Bloom filters



Probability of False Positive:

$$\left(1 - \left(1 - \frac{1}{m}\right)^{kn}\right)^k \approx \left(1 - e^{kn/m}\right)^k.$$

Fixed false positive -> m grows linearly with n

Implementation using Bloom filters

- Tokens
  - Produced each tick with probability p_{new}
  - Capped at g

- Registry
  - Big Bloom filter in the sky
    - Download and check if heard token is there
  - Can add through bit-wise OR
  - Plausible deniability

- Medical professional
  - User produces witness to infection
  - Uploads Bloom filter of infected to registry
  - Fake tokens won’t match what people hear

Implementation using Bloom filters – Sample Numbers

- Bloom filter
  - m=8x10^8
    - ~100MB / day download
    - No compression
    - No incremental updates
  - False Positive rate 10^{-15}
  - n<=11,000,000

- Town
  - 10,000 residents
  - ~1100 tokens / resident
  - 14 day history
    - 28,800 ticks @ 1 tick/minute

- 28800 / (1/ p_{new}) <= 1100
- p_{new} <= 3.5%
- recreate token every ~26 minutes

Analysis - Privacy

- User
  - Obtains tokens from (i) others, (ii) registry
  - Location information limited to epoch
  - Different epochs cannot be linked
- Registry
  - Cannot connect tokens, if updates batched
- Doctor
  - Can cause lots of damage
- World
  - Spoofing fake tokens
  - Rebroadcast others’ tokens

Philosophical problems

- Share your tokens
  - Bounded number of tokens
- Share heard tokens
  - Possible linkage (who else heard tokens)
- Shared encounter token
  - Complexity of interaction
- Reidentification
  - Few contacts
  - Cameras

Technical problems

- Bluetooth
  - Not all devices transmit at same power
  - Needs to be constantly receiving
    - problem on iOS
    - battery drain
- Token sharing
  - broadcast tokens are received tokens?
- Linkage
  - I know when I receive infected tokens
    - Reidentify sick person?

Extensions

- Bluetooth
  - Reduce power
  - Filter signal strength
    - RSSI
    - Packet loss

- Per-encounter tokens
  - A receives T_B from B
  - B receives T_A from A
  - Both compute $H(T_A,T_B)$

- Mediate server access
  - Requires trusted server
    - Allows monitoring access patterns
    - Use Private Set Intersection
  - Register tokens with callback

Extensions II

- Verify physical proximity
  - Multi-message handshake
  - Include coarselocation information in token

- Token
  - A receives T_B from B
  - B receives T_A from A
  - Both compute $H(T_A,T_B)$

- Staggered collocation
  - Devices in fixed locations

- Planned obsolescence
  - Data useless after infection window

Conclusion

- Adoption
  - Most important hurdle
  - Induce through fast-track testing
  - Induce through paying for positive conneciton
- Maintaining authenticity
  - Fake apps
  - Rogue apps
  - Patching apps
- Preventing abuse
  - Fake IDs

Acknowledgments

- Mayank Varia, co-author
- Ran Canetti, co-author
- Andy Sellars
- Gerald Denis
- Anand Devaiah
- Amir Herzberg
- David Starobinski
- Charles Write
- Ramesh Raskar
- Ron Rivest

Icons made by Freepik from www.flaticon.com