Decentralized Privacy-Preserving Proximity Tracing

UC San Diego
HALICIOĞLU DATA SCIENCE INSTITUTE

Haojian Jin
haojian@ucsd.edu

Before class quiz

1. What are the main differences between centralized and decentralized systems?
2. How many protocols does the paper introduce for the decentralized design? What are they?
3. What are the trade-offs of the three protocols? Which one did Apple/Android choose? Why?
4. Can you setup the code?
5. Teams?
6. Do you have a better design?

System design goals (Functionality, Privacy, Performance)

1. Coverage & accuracy
2. Authenticity
3. Security
4. Notification
5. Privacy
6. Performance
7. …

Why does privacy matter?

Exclusive: Government scientist Neil Ferguson resigns after breaking lockdown rules to meet his married lover

Prof Ferguson allowed the woman to visit him at home during the lockdown while lecturing the public on the need for strict social distancing

By Anna Mikhailova, DEPUTY POLITICAL EDITOR, Christopher Hope, CHIEF POLITICAL CORRESPONDENT, Michael Gillard *and* Louisa Wells

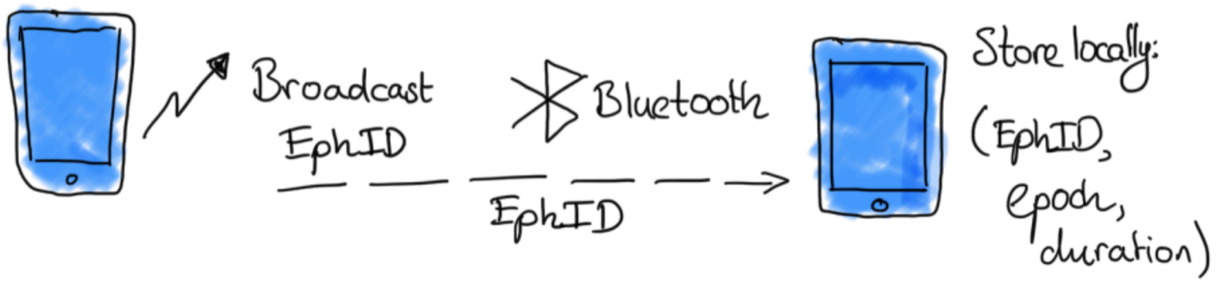
5 May 2020 • 7:17pm

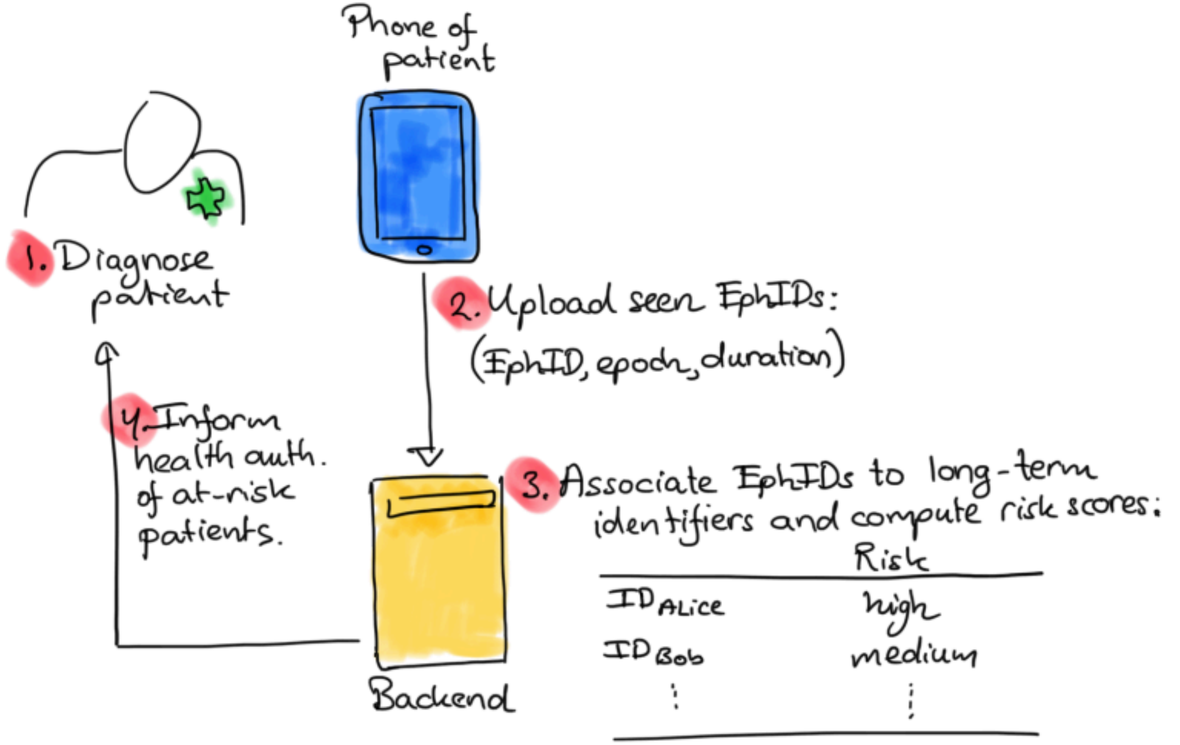
Neil Ferguson and Antonia Staats

The scientist whose advice prompted Boris Johnson to lock down Britain resigned from his Government advisory position on Tuesday night as The Telegraph can reveal he broke social distancing rules to meet his married lover....

1. Social graph
2. Interaction graph
3. Location traceability
4. At-risk individuals
5. Covid-19 positive status
6. Highly exposed locations

A centralized system





Why not centralized?

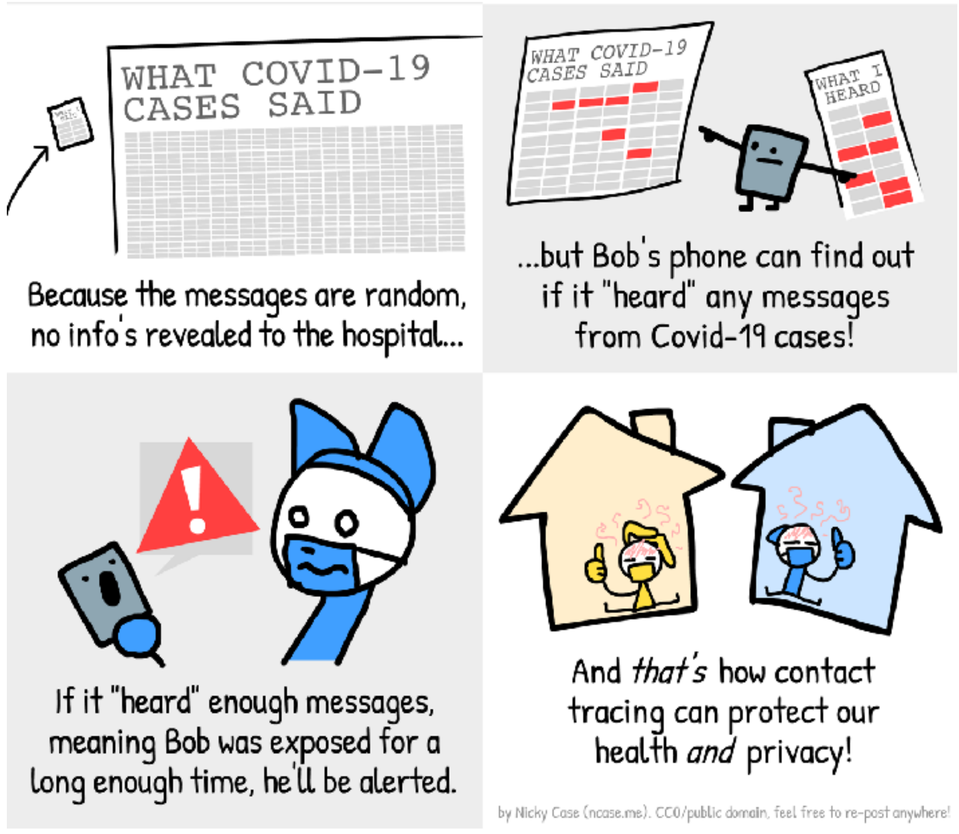
Resecurity
August 29, 2022

Share

COVID-19 data put for sale on the Dark Web

Resecurity, a California-based cybersecurity company protecting Fortune 500, has identified leaked PII stolen from Thailand's Department of Medical Sciences containing information about citizens with COVID-19 symptoms. The incident was uncovered and shared with Thai CERT.

How decentralized systems work?



https://www.youtube.com/watch?v=D__UaR5MQao

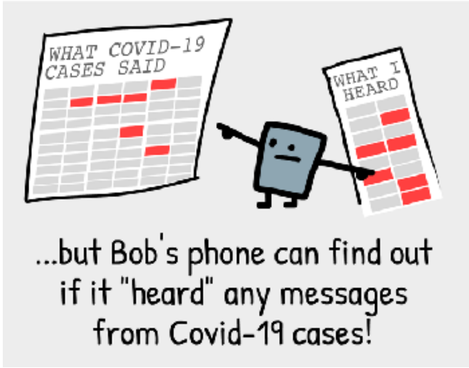
Design space

Alice's phone broadcasts a random message every few minutes.

How to generate the random messages?

If Alice gets Covid-19, she sends *her* messages to a hospital.

What data does Alice send to the server?



...but Bob's phone can find out if it "heard" any messages from Covid-19 cases!

What data does Bob receive from the server?

Crypto 101: Pseudorandom generator/function

`random.` **seed**(*a=None*, *version=2*)

Initialize the random number generator.

If *a* is omitted or `None`, the current system time is used. If randomness sources are provided by the operating system, they are used instead of the system time (see the `os.urandom()` function for details on availability).

If *a* is an int, it is used directly.

With version 2 (the default), a `str`, `bytes`, or `bytearray` object gets converted to an `int` and all of its bits are used.

With version 1 (provided for reproducing random sequences from older versions of Python), the algorithm for `str` and `bytes` generates a narrower range of seeds.

Changed in version 3.2: Moved to the version 2 scheme which uses all of the bits in a string seed.

Deprecated since version 3.9: In the future, the *seed* must be one of the following types: *NoneType*, `int`, `float`, `str`, `bytes`, or `bytearray`.

Deterministic

Random seed → Long string

(So hard to reverse engineering the seed.)

https://docs.python.org/3/library/random.html

Ephemeral ID generation

How to generate the random messages?

1. Each day, $SK_t = H(SK_{t-1})$

$$EphID_1 \,||\, \ldots \,||\, EphID_n = PRG(PRF(SK_t, \text{“broadcast key”}))$$

Pick a random order to transmit them, the phone stores SK_t it generated during the past 14 days.

Linkable

2. Each epoch draws a random 32-byte per-epoch seed

$$EphID_i = LEFTMOST128(H(seed_i)),$$

Non-linkable

3. At each time window, draws a random 16-byte per-epoch seed

$$EphID_{w,1} \,||\, \ldots \,||\, EphID_{w,n} = PRG(PRF(seed_w, \text{“DP3T-HYBRID”}))$$

Pick a random order to transmit them, the phone stores SK_t it generated during the past 14 days.

Hybrid
temporarily-linkable

Communication

If Alice gets Covid-19, she sends *her* messages to a hospital.

What data does Alice send to the server?

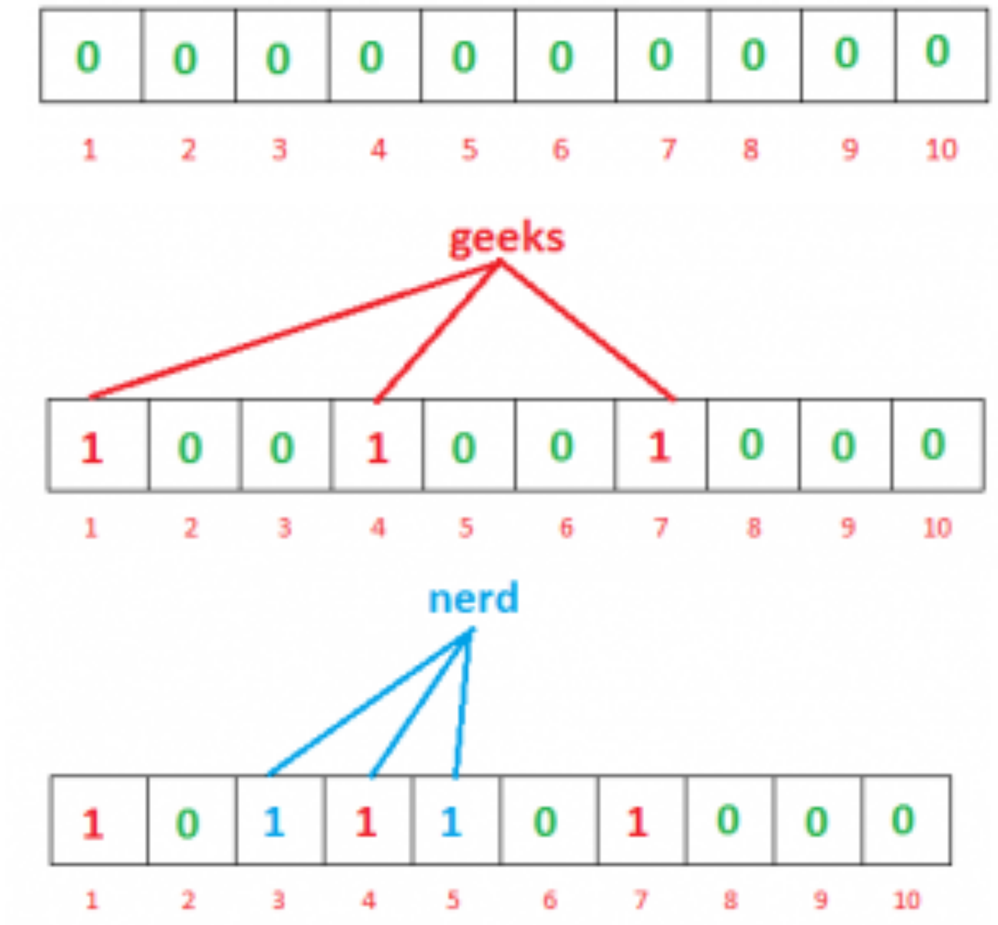
1. The backend collects the pairs (SK_t, t) of COVID-19 positive users.

2. Users have more fined control over their data. The users upload selected {i, Seed_i} to the server.

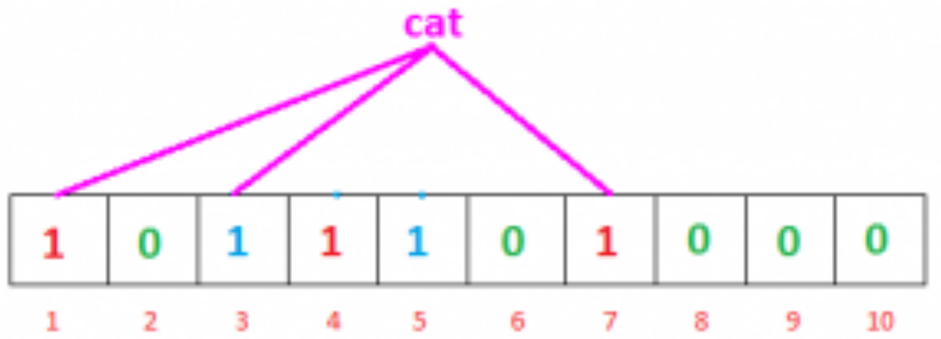
3. Users have more fined control over their data. The users upload selected {w, Seed_i} to the server.

Bloomfilter 101:

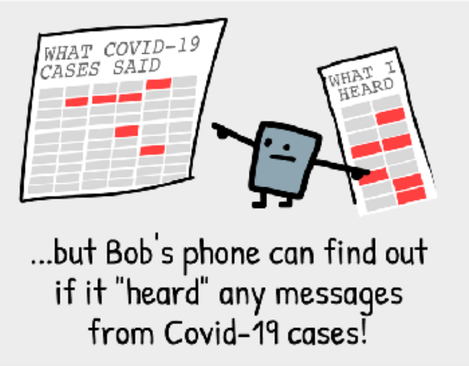
Insert



Lookup



Lookup



What data does Bob receive from the server?

1. Phones periodically download these pairs (SK_t, t) and use these pairs to reconstruct the list ephIDs.

2. The server maintains a public Cuckoo filter. Each smartphone uses the filter $\mathbb{F}$ to check.

Non-zero false positive!

3. Phones periodically download these pairs (w, seed_w) and use these pairs to reconstruct the list of EphIDs for each time window.