Role of Interaction Design & Data Science in Crisis Response

Karri-Pekka Laakso

Lead Designer (Interaction), Reaktor

Antti Honkela

Associate Professor at the Department of Computer Science, University of Helsinki and Coordinator of Research Programme in Privacy-preserving and Secure AI, Finnish Center for Artificial Intelligence (FCAI)

June 11, 2020

CS-E4002: Human-Centred Research and Design in Crisis

Aalto University

Ketju

Designing the Ketju Contact Tracing App: Interaction Design, Technology and Privacy Implications

11.6.2020

Karri-Pekka Laakso, Reaktor

Reaktor futurice Columbia Road FRAKTAL

In cooperation with

SITRA Vasa centralsjukhus Vaasan keskussairaala 2M IT

ketjusovellus.fi

- Officially started on Mon 23rd March, 2020
- 6 designers
    - 3 from Reaktor (UX, graphic, **UX**) <= me
    - 3 from Futurice (UX, graphic, UX)

Karri-Pekka Laakso

M.Sc. (eng), Helsinki University of Technology

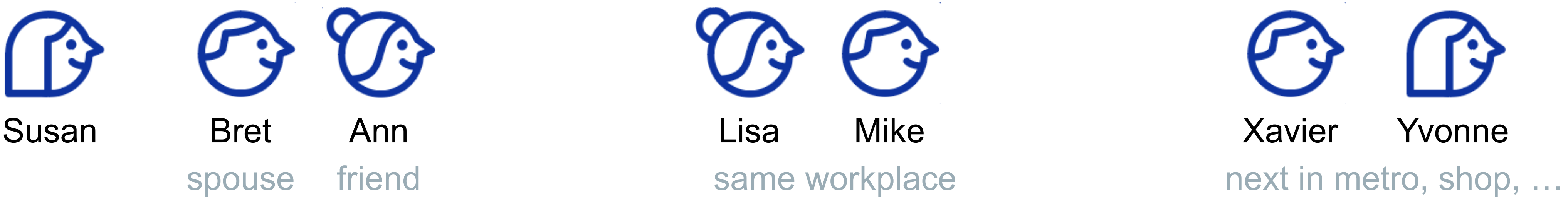
#UX #interaction_design #field_studies #visualizations

#scala #js #java #html_css #node #bash

#public_speaking #teaching

#family #gymnastics #acrobatics #orienteering #piano

Social Distance



Susan

Bret
spouse

Ann
friend

Lisa
Mike
same workplace

Xavier
Yvonne
next in metro, shop, …

Who can Susan recall / name?

Susan

Bret
spouse

Ann
friend

Lisa Mike
same workplace

Xavier Yvonne
next in metro, shop, …

Who would Susan share with?

Susan

Bret
spouse

Ann
friend

Lisa Mike
same workplace

Xavier Yvonne
next in metro, shop, …

Who knows Susan by name?

Susan

Bret Ann

spouse friend

Lisa Mike

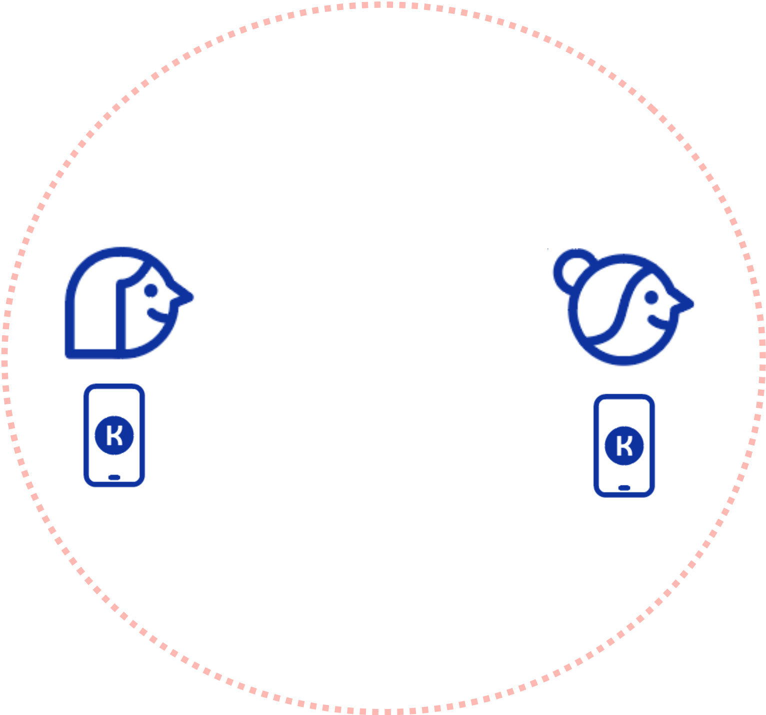
same workplace

Xavier Yvonne

next in metro, shop, …

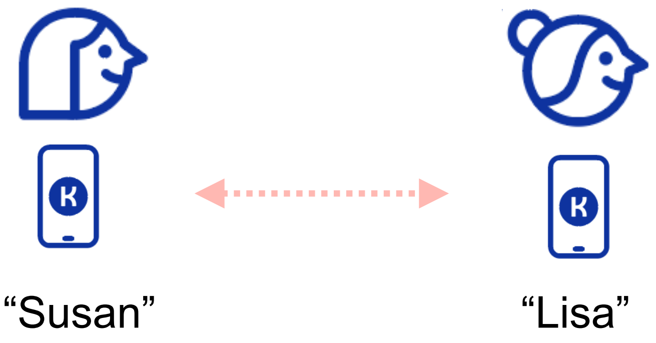
Technology and privacy

- Location



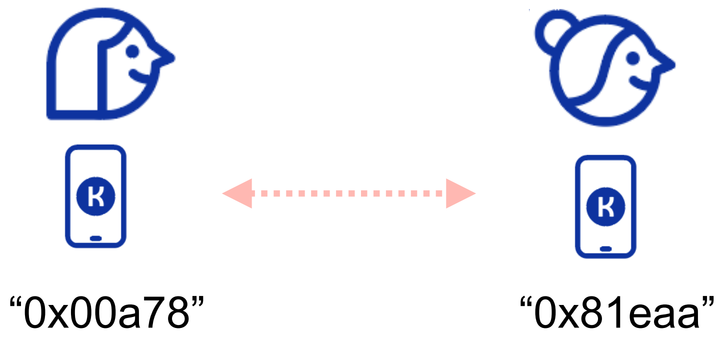
Technology and privacy

- ~~Location~~
- Bluetooth contacts



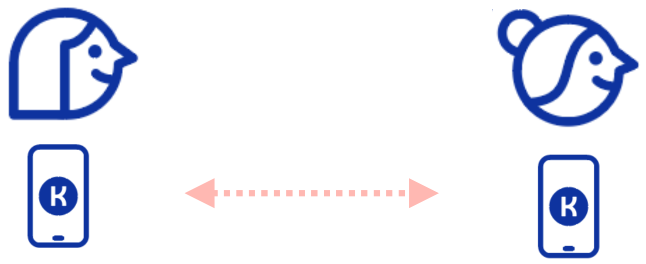
Technology and privacy

- ~~Location~~
- ~~Bluetooth contacts~~
- Anonymous Bluetooth contacts



Technology and privacy

- ~~Location~~
- ~~Bluetooth contacts~~
- ~~Anonymous Bluetooth contacts~~
- Anonymous Bluetooth contacts changing in time (DP^3T, Google Apple Exposure Notification API)



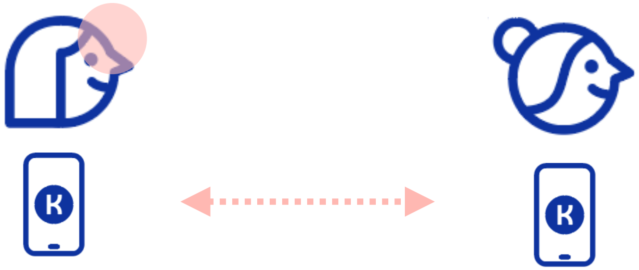
15 min	“0x00a78”	“0x81eaa”
15 min	“0xb2312”	“0xa292f”
15 min	“0x9124a”	“0x93d51”
	…	…

Technology and privacy

“0x00a78”

“0xb2312”

“0x9124a”



I have met

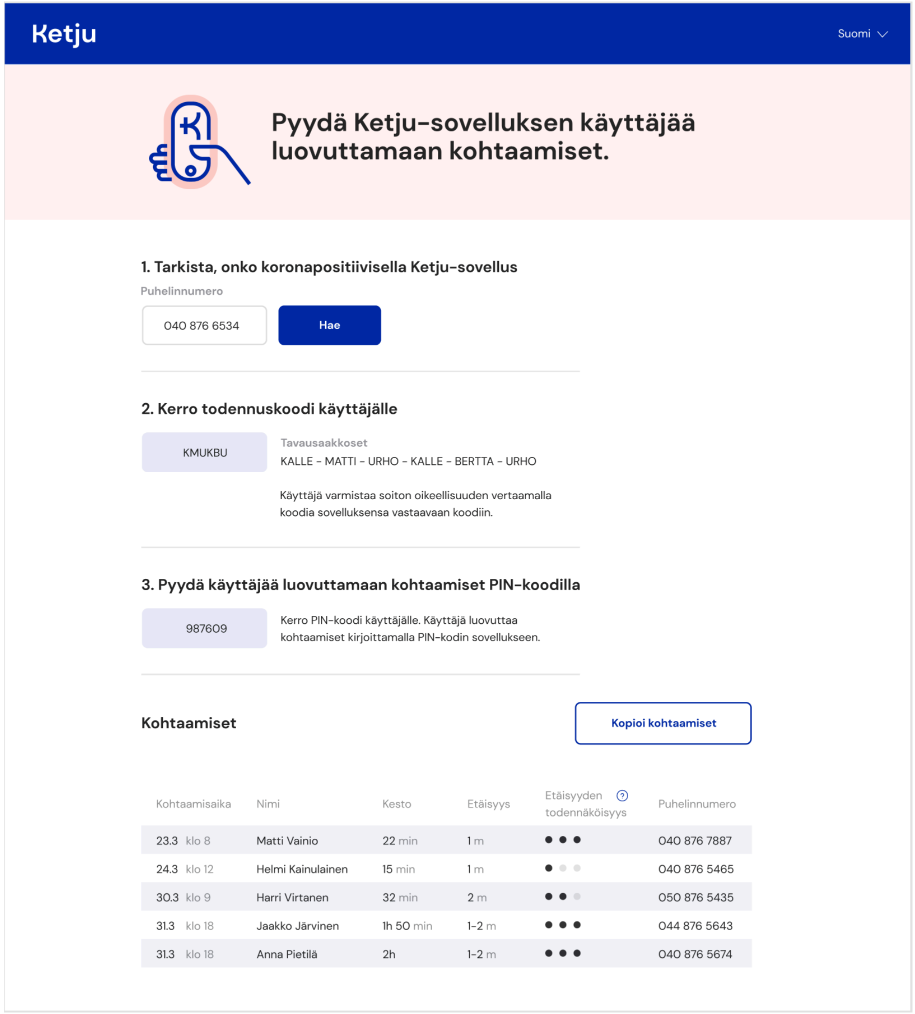
9:05 - 9:12 “0x00a78”

9:12 - 9:17 “0xb2312”

Centralized model

- Authorities know the exposed immediately
- Gather names & phone numbers => call, interview, quarantine

- Privacy?



Pyydä Ketju-sovelluksen käyttäjää luovuttamaan kohtaamiset.

1. Tarkista, onko koronapositiivisella Ketju-sovellus

Puhelinnumero: 040 876 6534

2. Kerro todennuskoodi käyttäjälle

KMUKBU

Tavausaakkoset
KALLE - MATTI - URHO - KALLE - BERTTA - URHO

Käyttäjä varmistaa soiton oikeellisuuden vertaamalla koodia sovelluksensa vastaavaan koodiin.

3. Pyydä käyttäjää luovuttamaan kohtaamiset PIN-koodilla

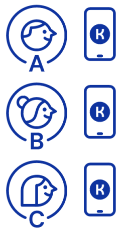
987609

Kerro PIN-koodi käyttäjälle. Käyttäjä luovuttaa kohtaamiset kirjoittamalla PIN-koodin sovellukseen.

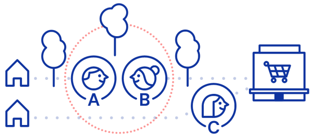
Kohtaamiset

Kohtaamisaika	Nimi	Kesto	Etäisyys	Etäisyyden todennäköisyys	Puhelinnumero
23.3 klo 8	Matti Vainio	22 min	1 m	● ● ●	040 876 7887
24.3 klo 12	Helmi Kainulainen	15 min	1 m	● ○ ○	040 876 5465
30.3 klo 9	Harri Virtanen	32 min	2 m	● ● ○	050 876 5435
31.3 klo 18	Jaakko Järvinen	1h 50 min	1–2 m	● ● ●	044 876 5643
31.3 klo 18	Anna Pietilä	2h	1–2 m	● ● ●	040 876 5674

Distributed model



1) Citizens load the Ketju app from the App Store and put it on every time they leave home.

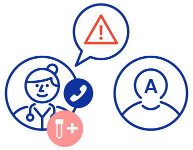


2) When users A and B meet, their phones register a contact when are close enough for time long enough (for example <5 m, >15 min).

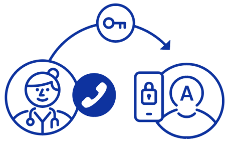
C is too far away to register a contact.

3) The app stores the contact totally anonymously: no personal or location information is stored.

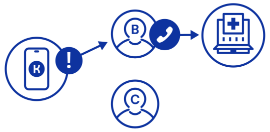
4) A has symptoms and the doctor order a test.



5) The doctor calls to tell that the test result is positive and asks, if A uses the Ketju App.



6) The doctor gives A a PIN code, which A enters to the Ketju app. The app informs the central server that A has been infected.



7) The app of B notices that one of its contacts has reported a positive test result. The app tells B about the exposure and that she should contact the local health center.

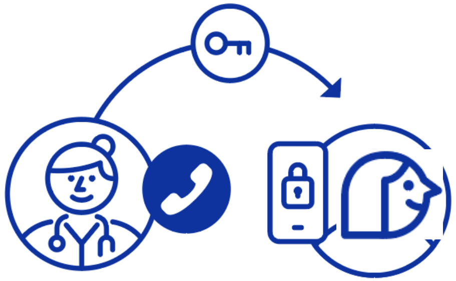
C has had no contact with A and thus nothing happens.

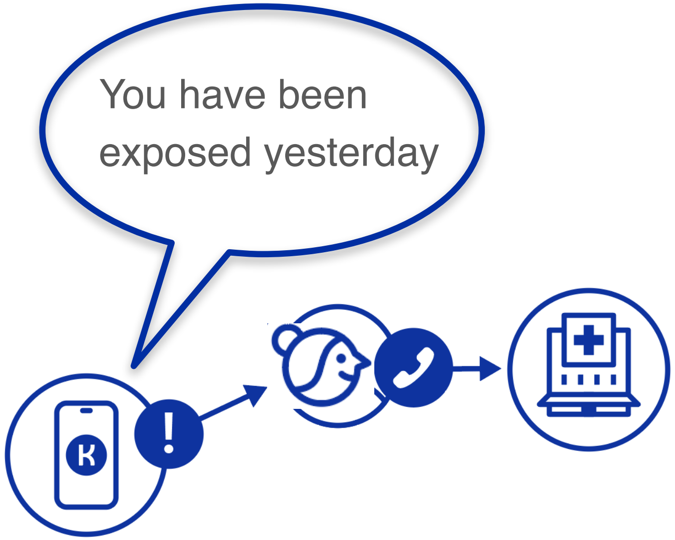
8) Only the user knows about the possible infection, since the data at the server and at the phone in anonymous.

Distributed model

- Susan tells the system that she is sick (voluntary)
- Lisa gets a notification
  => take a test (voluntary)
  => self-quarantine (voluntary)

- Verification of exposure?
- Granularity of exposure time?
- Contact tracing?

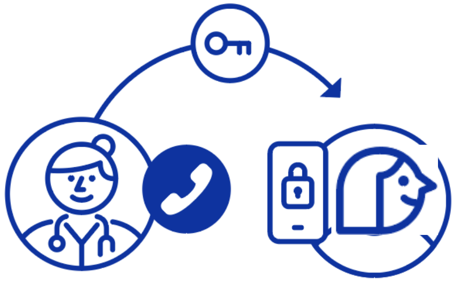




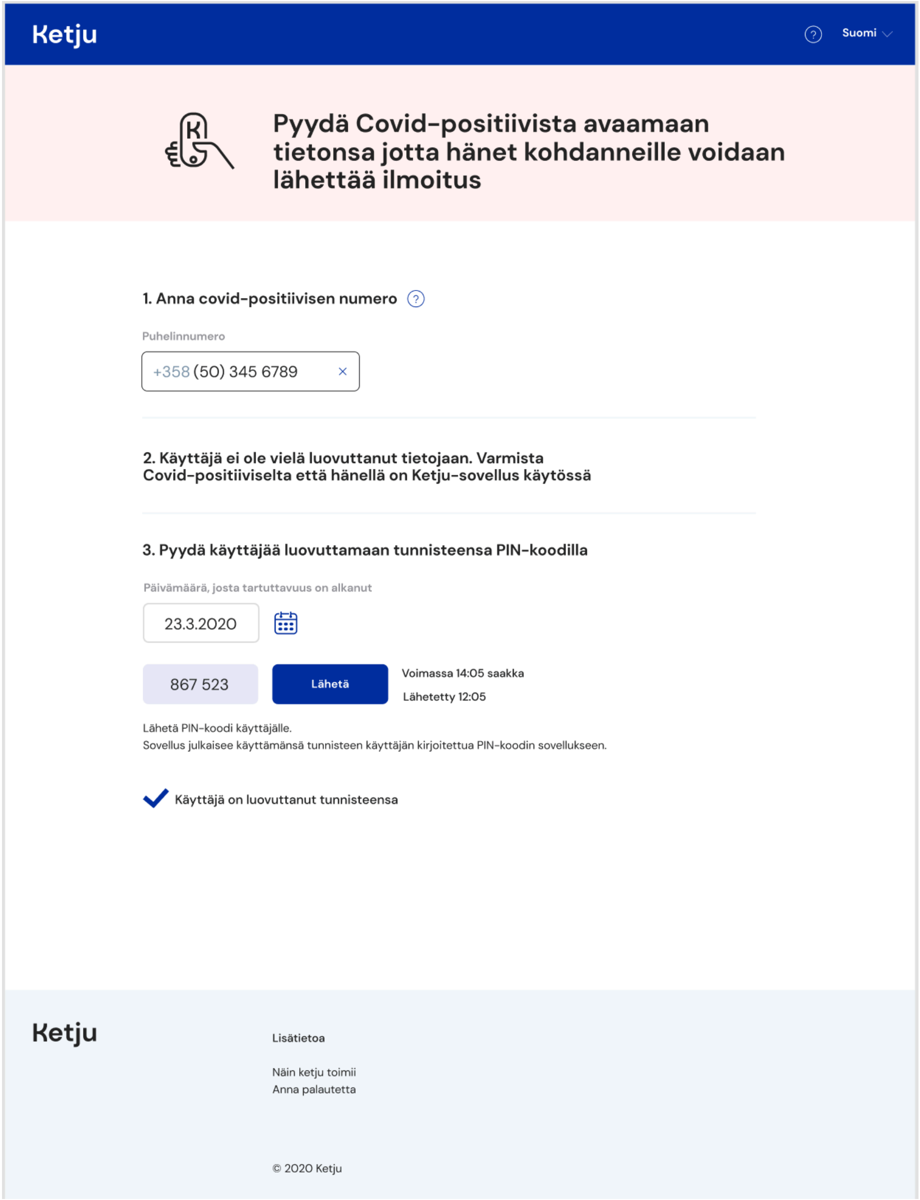
Distributed model

Authorities can only give the release code

=> no help for contact tracing
=> no official quarantine
(≠ self quanrantine)



Ketju



Pyydä Covid-positiivista avaamaan tietonsa jotta hänet kohdanneille voidaan lähettää ilmoitus

1. Anna covid-positiivisen numero

Puhelinnumero: +358 (50) 345 6789

2. Käyttäjä ei ole vielä luovuttanut tietojaan. Varmista Covid-positiiviselta että hänellä on Ketju-sovellus käytössä

3. Pyydä käyttäjää luovuttamaan tunnisteensa PIN-koodilla

Päivämäärä, josta tartuttavuus on alkanut: 23.3.2020

867 523

Voimassa 14:05 saakka
Lähetetty 12:05

Lähetä PIN-koodi käyttäjälle.
Sovellus julkaisee käyttämänsä tunnisteen käyttäjän kirjoitettua PIN-koodin sovellukseen.

✓ Käyttäjä on luovuttanut tunnisteensa

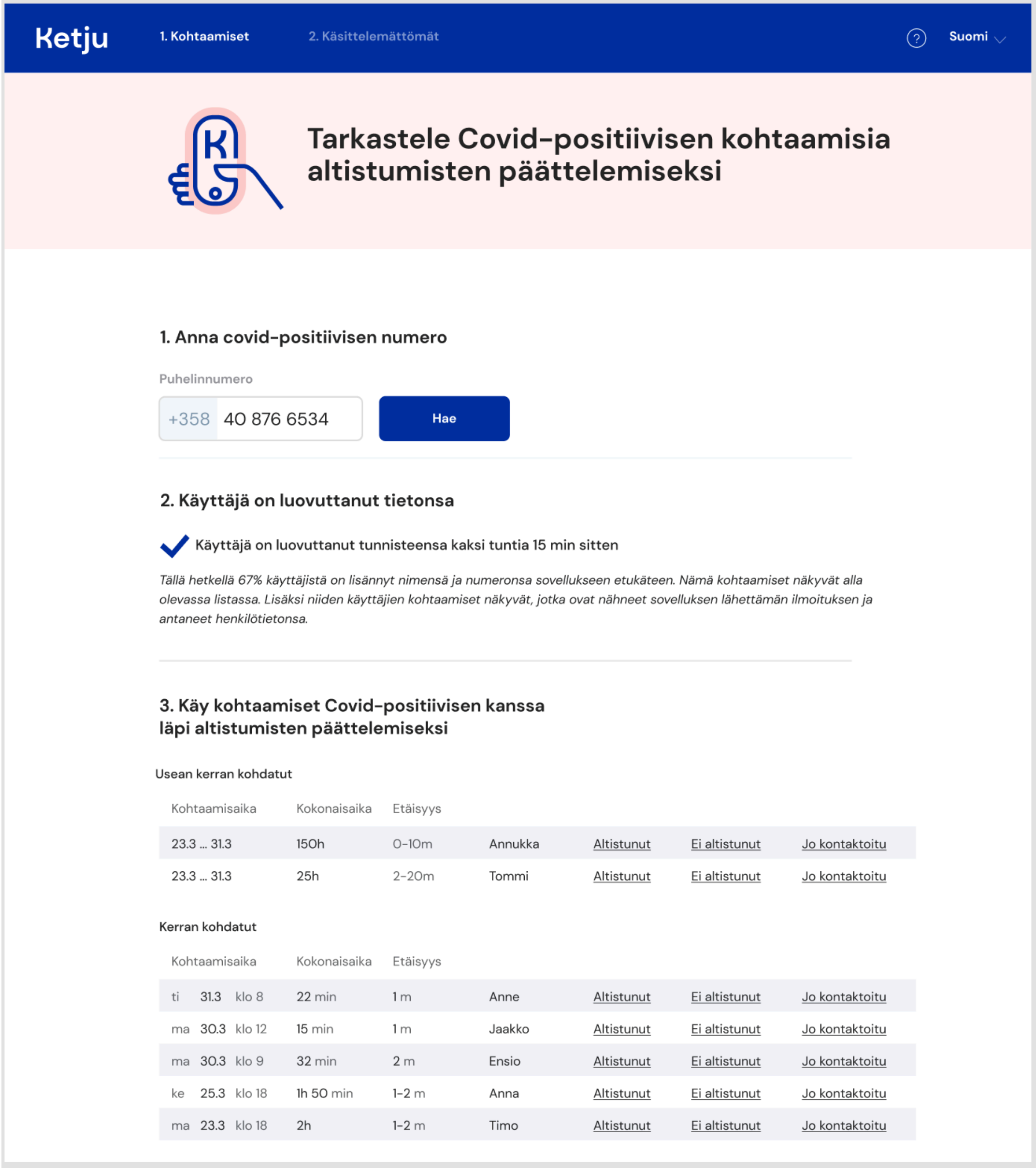
© 2020 Ketju

Hybrid models

- People voluntarily release information to authorities, e.g.
  - contacts with a sick person
  - exposures of a sick person
  - their contact info
  - …

- Legal issues
- Privacy issues
- Gapple API issues

Ketju



Tarkastele Covid-positiivisen kohtaamisia altistumisten päättelemiseksi

1. Anna covid-positiivisen numero

Puhelinnumero

+358 40 876 6534

2. Käyttäjä on luovuttanut tietonsa

✓ Käyttäjä on luovuttanut tunnisteensa kaksi tuntia 15 min sitten

Tällä hetkellä 67% käyttäjistä on lisännyt nimensä ja numeronsa sovellukseen etukäteen. Nämä kohtaamiset näkyvät alla olevassa listassa. Lisäksi niiden käyttäjien kohtaamiset näkyvät, jotka ovat nähneet sovelluksen lähettämän ilmoituksen ja antaneet henkilötietonsa.

3. Käy kohtaamiset Covid-positiivisen kanssa läpi altistumisten päättelemiseksi

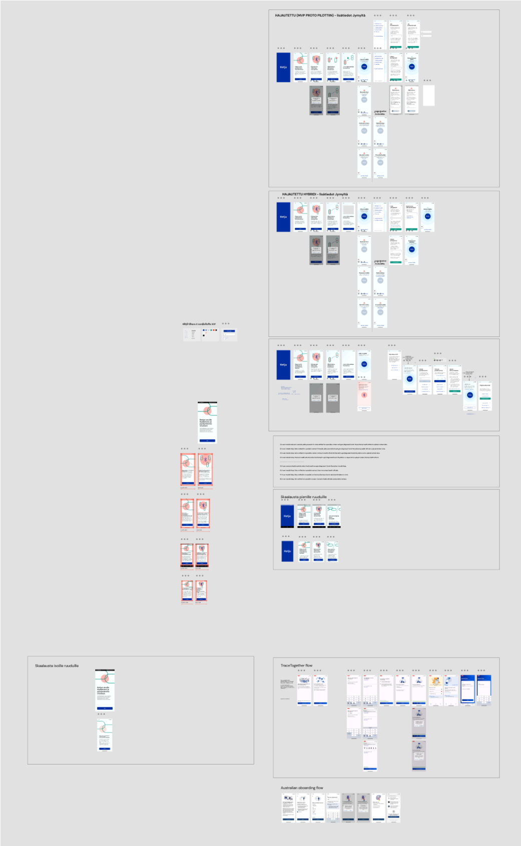
Usean kerran kohdatut

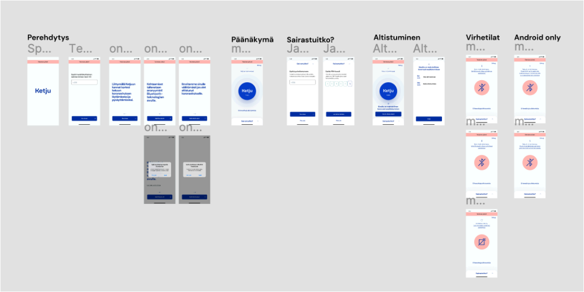
Kohtaamisaika	Kokonaisaika	Etäisyys				
23.3 … 31.3	150h	0–10m	Annukka	Altistunut	Ei altistunut	Jo kontaktoitu
23.3 … 31.3	25h	2–20m	Tommi	Altistunut	Ei altistunut	Jo kontaktoitu

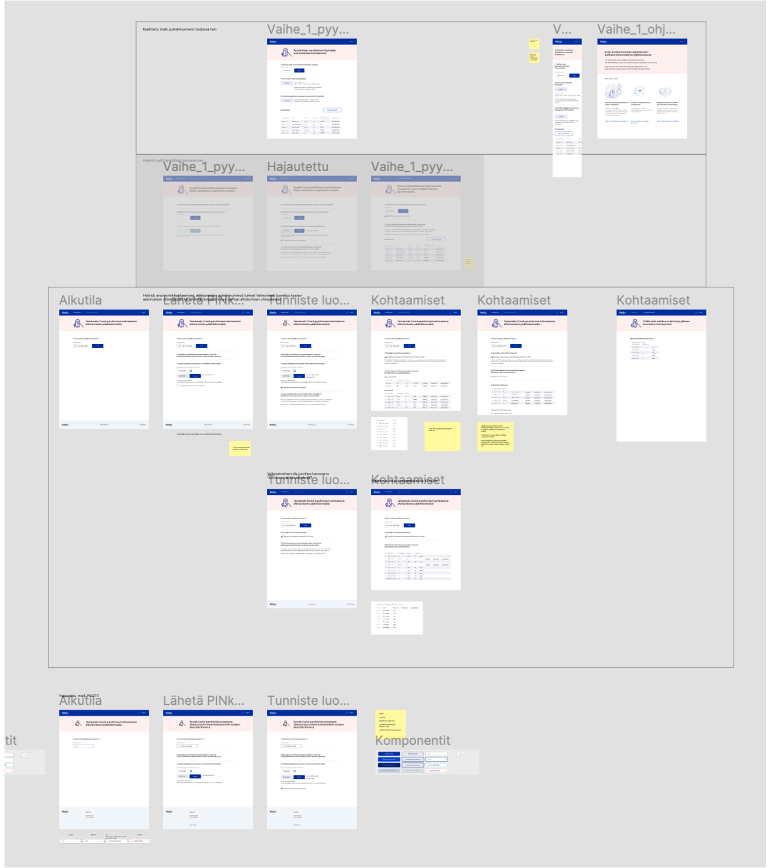
Kerran kohdatut

Kohtaamisaika	Kokonaisaika	Etäisyys				
ti 31.3 klo 8	22 min	1 m	Anne	Altistunut	Ei altistunut	Jo kontaktoitu
ma 30.3 klo 12	15 min	1 m	Jaakko	Altistunut	Ei altistunut	Jo kontaktoitu
ma 30.3 klo 9	32 min	2 m	Ensio	Altistunut	Ei altistunut	Jo kontaktoitu
ke 25.3 klo 18	1h 50 min	1–2 m	Anna	Altistunut	Ei altistunut	Jo kontaktoitu
ma 23.3 klo 18	2h	1–2 m	Timo	Altistunut	Ei altistunut	Jo kontaktoitu

Iterations







Ketju

Chaotic surroundings

contract tracing

doctors

politics

infectious diseases

communications

law

media

design

users

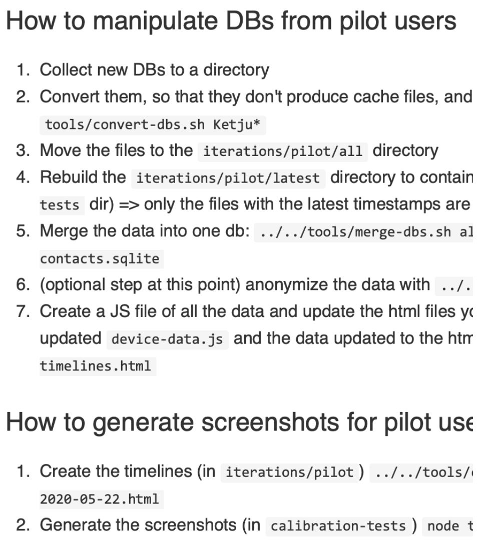
bluetooth

protocols

databases

visualizations

Do what is needed

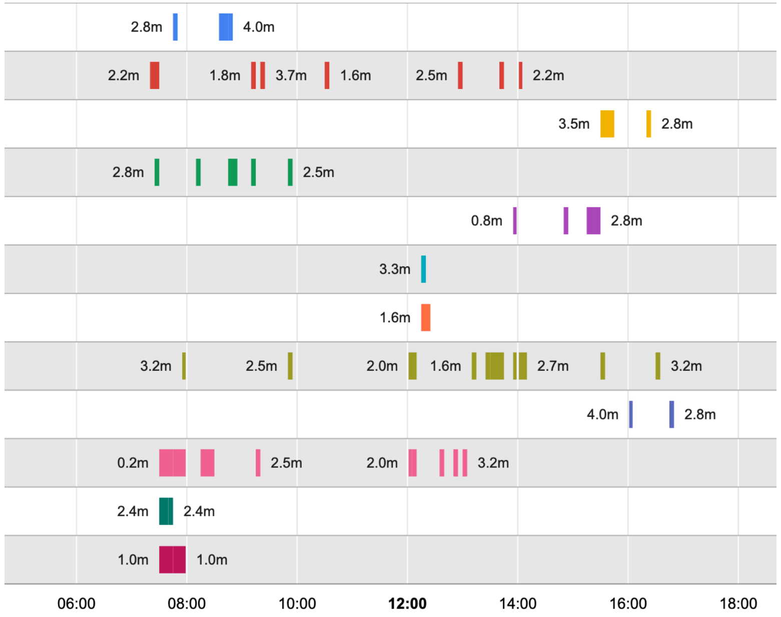


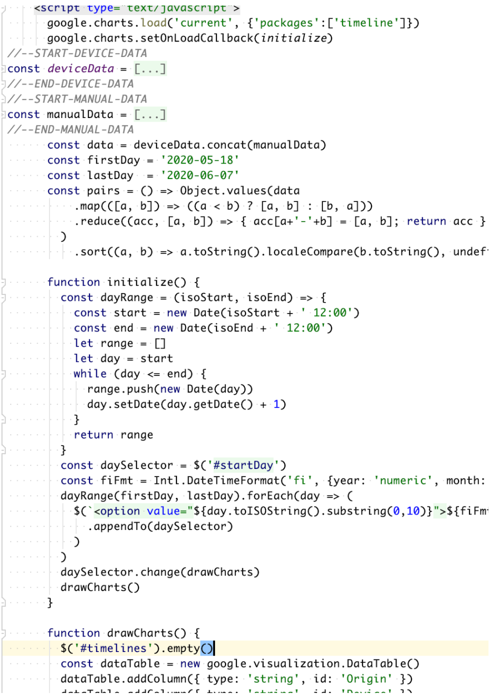
How to manipulate DBs from pilot users

1. Collect new DBs to a directory
2. Convert them, so that they don't produce cache files, and
   `tools/convert-dbs.sh Ketju*`
3. Move the files to the `iterations/pilot/all` directory
4. Rebuild the `iterations/pilot/latest` directory to contain
   `tests` dir) => only the files with the latest timestamps are
5. Merge the data into one db: `../../tools/merge-dbs.sh al`
   `contacts.sqlite`
6. (optional step at this point) anonymize the data with `../.`
7. Create a JS file of all the data and update the html files yo
   updated `device-data.js` and the data updated to the htm
   `timelines.html`

How to generate screenshots for pilot use

1. Create the timelines (in `iterations/pilot`) `../../tools/`
   `2020-05-22.html`
2. Generate the screenshots (in `calibration-tests`) `node t`





```
  <script type="text/javascript">
    google.charts.load('current', {'packages':['timeline']})
    google.charts.setOnLoadCallback(initialize)
//--START-DEVICE-DATA
const deviceData = [...]
//--END-DEVICE-DATA
//--START-MANUAL-DATA
const manualData = [...]
//--END-MANUAL-DATA
      const data = deviceData.concat(manualData)
      const firstDay = '2020-05-18'
      const lastDay  = '2020-06-07'
      const pairs = () => Object.values(data
          .map(([a, b]) => ((a < b) ? [a, b] : [b, a]))
          .reduce((acc, [a, b]) => { acc[a+'-'+b] = [a, b]; return acc }
        )
          .sort((a, b) => a.toString().localeCompare(b.toString(), undef

      function initialize() {
        const dayRange = (isoStart, isoEnd) => {
          const start = new Date(isoStart + ' 12:00')
          const end = new Date(isoEnd + ' 12:00')
          let range = []
          let day = start
          while (day <= end) {
            range.push(new Date(day))
            day.setDate(day.getDate() + 1)
          }
          return range
        }
        const daySelector = $('#startDay')
        const fiFmt = Intl.DateTimeFormat('fi', {year: 'numeric', month:
        dayRange(firstDay, lastDay).forEach(day => (
          $(`<option value="${day.toISOString().substring(0,10)}">${fiFm
            .appendTo(daySelector)
          )
        )
        daySelector.change(drawCharts)
        drawCharts()
      }

      function drawCharts() {
        $('#timelines').empty()
        const dataTable = new google.visualization.DataTable()
        dataTable.addColumn({ type: 'string', id: 'Origin' })
```

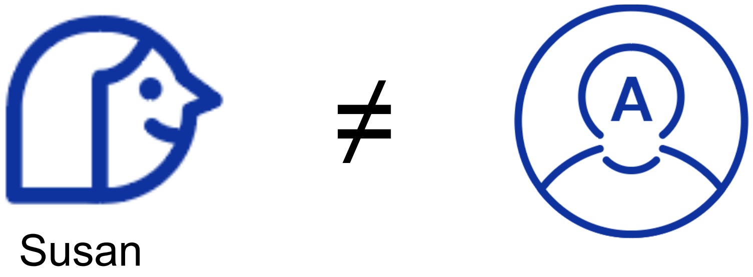
produce 330 screenshots for users

organize 250 database files

participate in writing the report

Ketju

The power of realistic cases



Everything is easier: talking, deducing, …

Cost / benefit

x% install
* y% use
* z% contact noticed
* i% user notices
* j% contacts the authority
* …

vs. cost of implementation & maintenance?

Ketju

Questions?

karri@reaktor.fi

@kplaakso

Privacy-preserving contact statistics collection using COVID-19 contact tracing apps

Antti Honkela[1] and Tejas Kulkarni[2]

Finnish Center for Artificial Intelligence FCAI
[1] University of Helsinki
[2] Aalto University

Human-Centred Research and Design in Crisis
11 June 2020

FCAI Finnish
Center for
Artificial
Intelligence

Outline

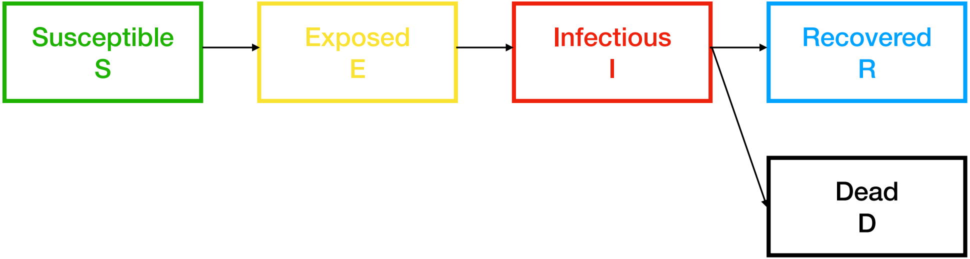
A cartoon of an epidemic

Contact tracing apps

Privacy-preserving collection of contact statistics

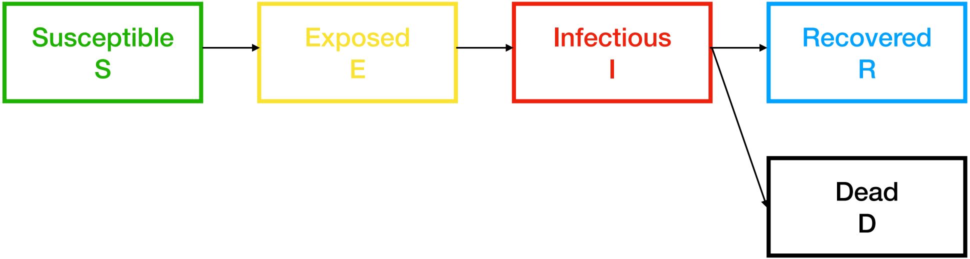
Disclaimer: while I talk about mathematical epidemic models, I am not an epidemiologist. Listeners beware.

A cartoon of an epidemic



- Population divided to classes of individuals based on infection status
- Infection spreads when S has sufficiently strong contact with I
- Current growth rate of the epidemic measured by reproductive number $R_e \approx$ ratio of new infections over recoveries

A cartoon of an epidemic



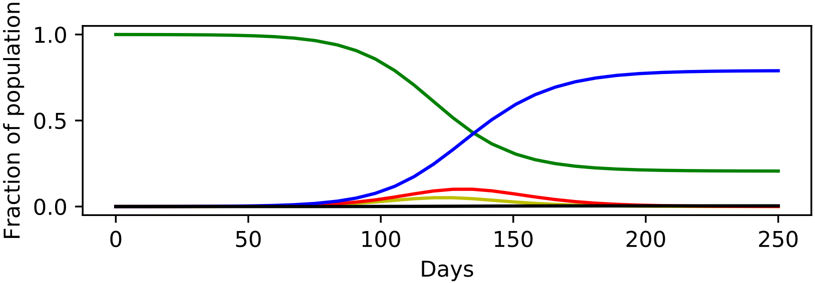
Susceptible
S

Exposed
E

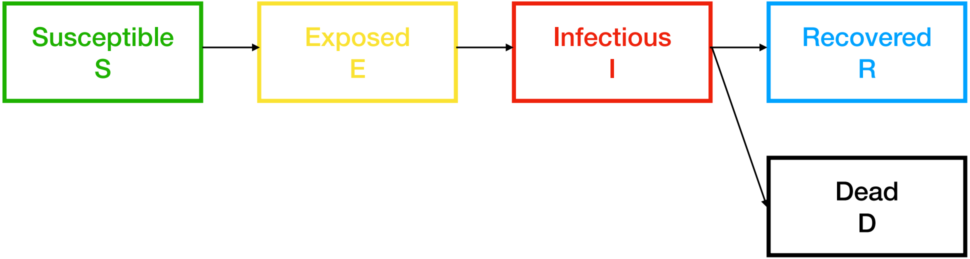
Infectious
I

Recovered
R

Dead
D

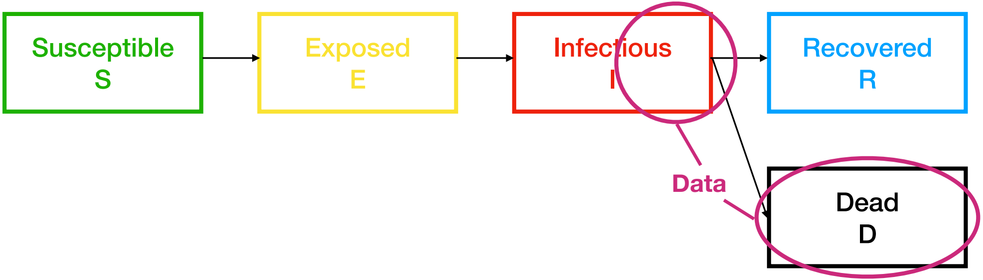


A cartoon of an epidemic



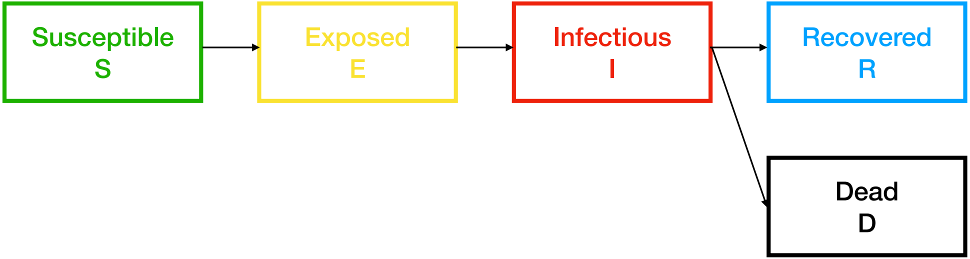
- Modelling by fitting the curve to observed confirmed cases / hospitalisations / deaths / ...
- Estimation delay: changes in infection rate only show in tests and hospitalisations after a week or more

A cartoon of an epidemic



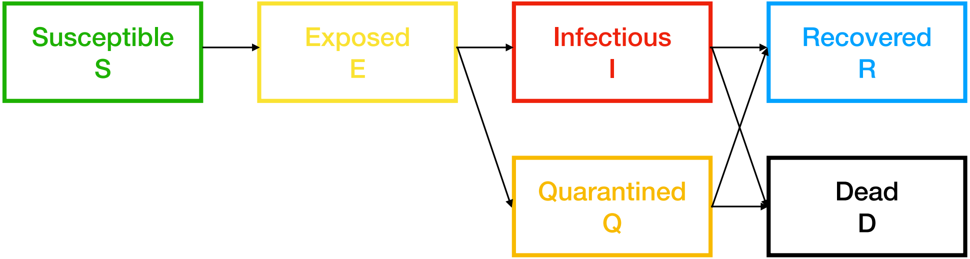
- Modelling by fitting the curve to observed confirmed cases / hospitalisations / deaths / ...
- Estimation delay: changes in infection rate only show in tests and hospitalisations after a week or more

Managing an epidemic



- Epidemic management activities aim at limiting contacts between S and I
  - Social distancing limits all contacts
  - Contact tracing aims to quarantine exposed individuals before they become infective

Managing an epidemic

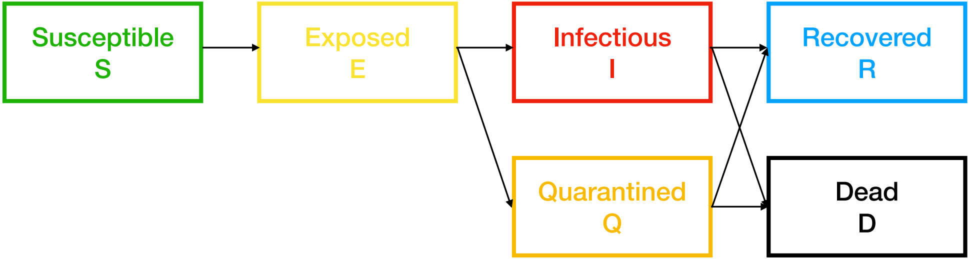


- Epidemic management activities aim at limiting contacts between S and I
  - Social distancing limits all contacts
  - Contact tracing aims to quarantine exposed individuals before they become infective

Contact tracing and mobile apps

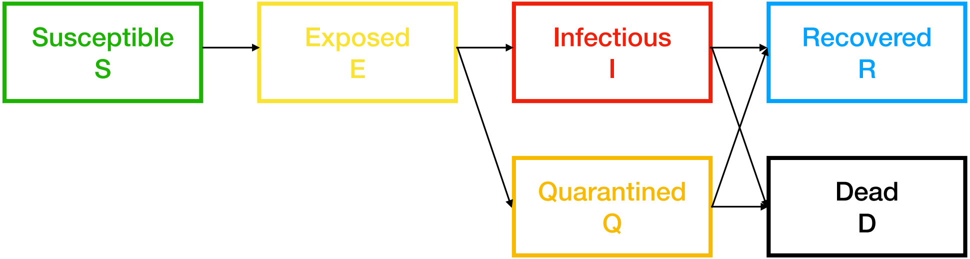
- Classical contact tracing works remarkably well, but is very labour intensive
- Many countries are developing mobile apps to assist the activity
- Privacy is an absolute requirement
- Typical mode of operation (e.g. DP-3T, Google/Apple):
  - No location data gathered (low utility, bad privacy)
  - Use Bluetooth to record random identifiers broadcast by nearby devices
  - Identifiers change relatively frequently
  - Recorded recent contacts shared only when user is diagnosed positive
- Discussion to decide between decentralised vs. centralised mode of operation

Contact tracing app theory



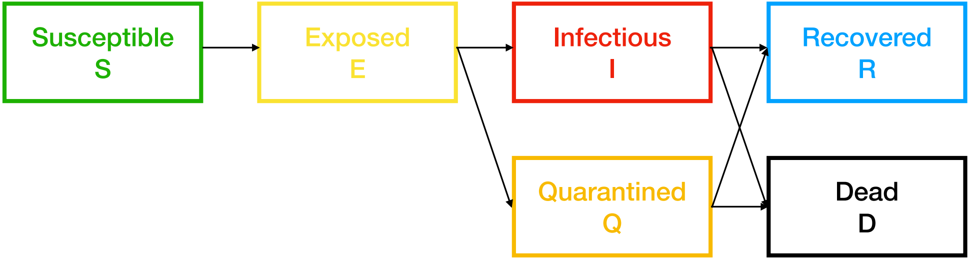
- Assuming 50% of population use the app and all detected exposed individuals are perfectly quarantined, we prevent 25% of potential future infections

Contact tracing app theory



- Assuming 50% of population use the app and all detected exposed individuals are perfectly quarantined, we prevent 25% of potential future infections
  - E.g. reduce $R_e = 1.2 \rightarrow R_e = 0.9$

Contact tracing app theory



- Assuming 50% of population use the app and all detected exposed individuals are perfectly quarantined, we prevent 25% of potential future infections
  - E.g. reduce $R_e = 1.2 \rightarrow R_e = 0.9$
- Over 70% users needed for theoretical 50% efficiency
  - E.g. reduce $R_e = 1.8 \rightarrow R_e = 0.9$

Opportunity

- Contact tracing apps collect data of close contacts in the population
- Collecting statistical information of contact frequencies would provide data for epidemic modelling as well as monitoring and planning other interventions (e.g. school and business closures)
- While such statistical information is not very sensitive, strong privacy protection is still necessary; collection should be opt-in
- Such data likely to be useful with fewer users: even at 30% use we would expect to see 30% of contacts of each user and can correct that from known rate of users

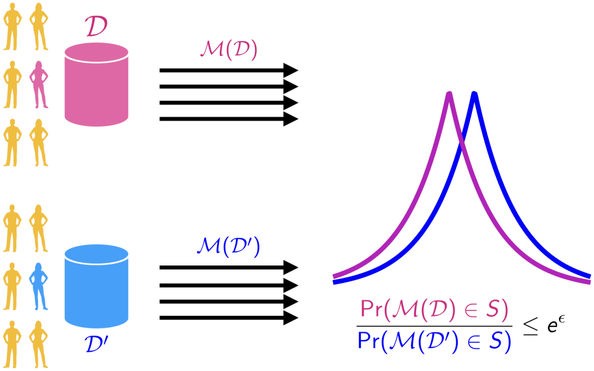
Asking sensitive questions: Randomised response (Warner, 1965)

Assume respondents are instructed to answer a potentially sensitive query (e.g. were you in contact with more than 10 individuals yesterday?) as follows:

1. Flip a coin in secret.
2. If **tails**, then respond truthfully.
3. If **heads**, then flip a second coin and respond “Yes” if heads and “No” if tails.

- Outcome: the answer is flipped with probability $\frac{1}{4}$
- Everyone gets plausible deniability: “It was just the coins”
- Statistics can be estimated from population responses by compensating for the noise
- Probabilistic loss of privacy: “Yes” response makes it more likely your true response was “Yes”

Differential privacy (DP; Dwork *et al.*, 2006)



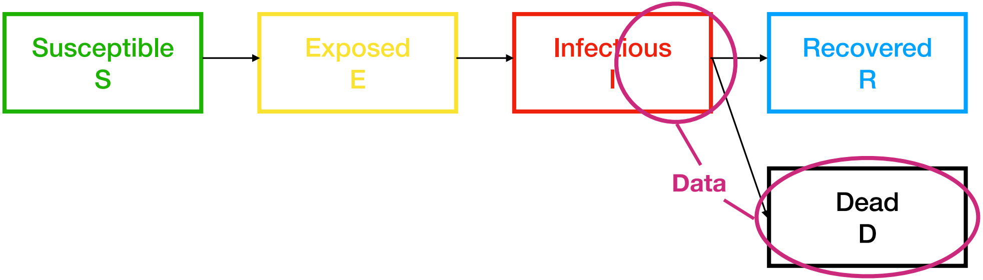
$$\frac{\Pr(\mathcal{M}(\mathcal{D}) \in S)}{\Pr(\mathcal{M}(\mathcal{D}') \in S)} \leq e^{\epsilon}$$

- Provides protection against adversaries with side information
- Degrades gracefully under repeated use
- Invariant to post-processing

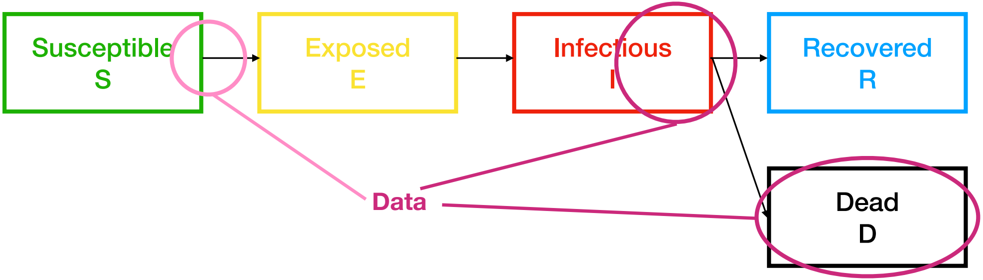
Proposal: Locally DP collection of contact statistics

- Using local DP (LDP) to collect population histogram of number of contacts
- Data are anonymised before they leave the user device
- Privacy guarantee: even if you see a user's report knowing it came from her, you can only guess what she answered
- Rapidly changing identifiers make collection of full daily statistics difficult
  - Collect e.g. maximum number of contacts over 30 min period each day instead
- Observed data fed to a Bayesian model for denoising, integrating responses from multiple days in a probabilistic model
  - Produces real-time data on behaviour changes and effects of changes to interventions and guidelines

Extended data collection



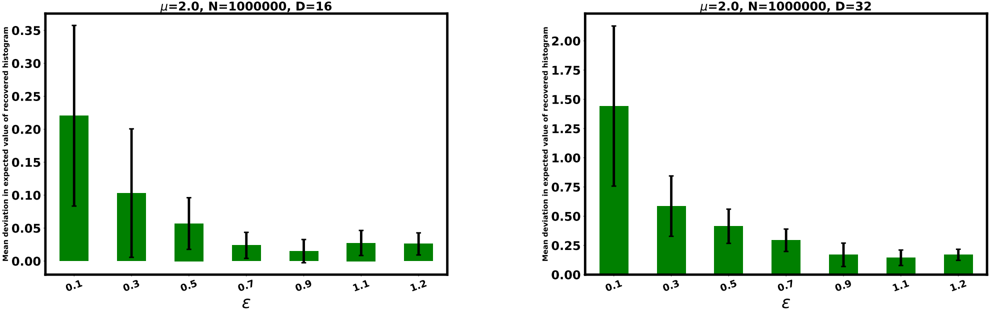
Extended data collection



More about privacy

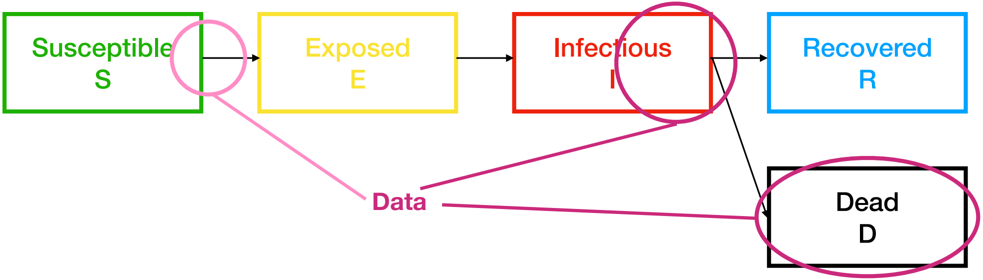
- Caveat: theoretical privacy guarantees will degrade over repeated data collection
  - Using fresh data for each report mitigates the impact
  - Ethical review of costs vs. benefits still needed
- Introducing *secure shuffling* to eliminate linking reports to individuals tightens the privacy guarantee by approximately factor $\sqrt{n}$ for n users
  - Basically eliminates privacy concerns even after repeated collection
  - NB: even breaking the shuffler would not completely compromise the privacy

Preliminary simulation results



Error in estimating the mean of a simulated population using LDP histogram estimation with D bins. Simulated using geometric distribution with true mean 2.0. Bars show MAE of 10 repeats, error bars show std over the repeats.

Conclusion



- COVID-19 pandemic has just started, still long way to vaccine or herd immunity
- Daily contact statistics observed by contact tracing apps could provide direct measures of the most important spreading mechanism, enabling more timely modelling
- Differential privacy provides the means to collect these data under strong privacy guarantees to the users

Next Tuesday:

Why Contact Tracing is Neither Secure or Privacy-Preserving

Janne Lindqvist,

Prof. at the Department of Computer Science, Aalto University

Aalto University
School of Science