Mobile Sensing: Privacy Limits and Ethics

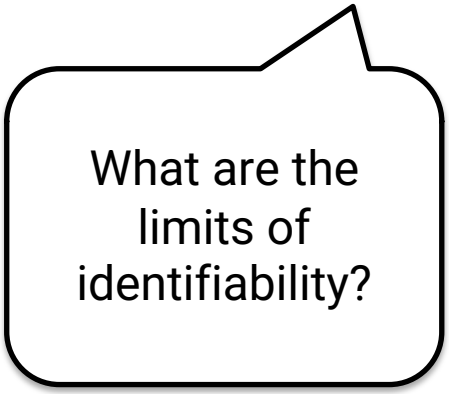
Master studies, 2021/2022

Dr Veljko Pejović
Veljko.Pejovic@fri.uni-lj.si

University *of Ljubljana*
Faculty *of Computer and Information Science*

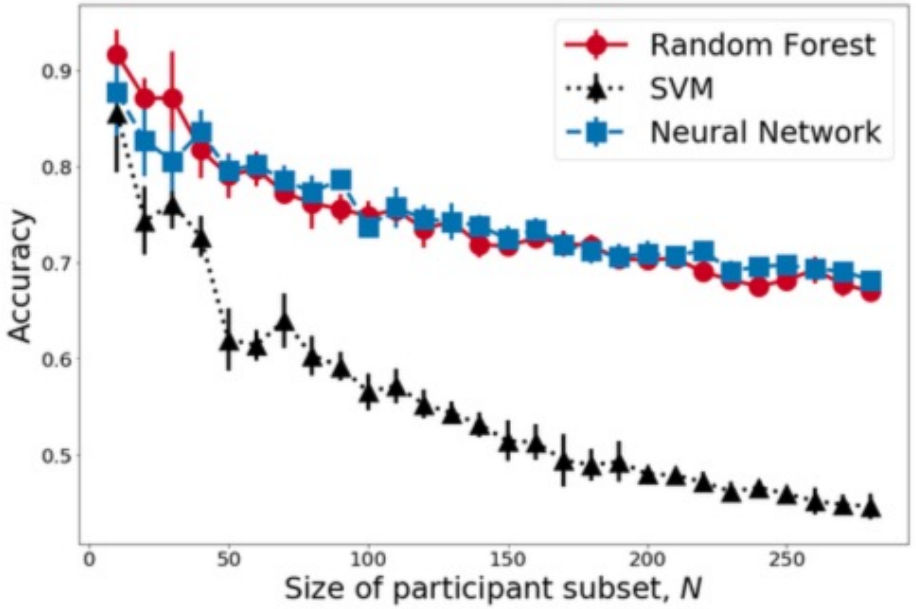
Sensor Data-Based Identification

- Biometrics
  - Fingerprints, face photos, voice properties, iris scans, heartbeats, etc. can be used to identify a person
- Behavioral properties
  - Gait, keyboard typing styles, locations visited, applications installed, etc. can be used to identify a person



Limits of Identifiability

- More users in a dataset, more difficult it is to identify an individual (in general)



If you have a low number of users even very "benign" data can be identifiable

"Recruit Until It Fails: Exploring Performance Limits for Identification Systems", Sugrim et al, ACM IMWUT 2019

Mobility as Identifier

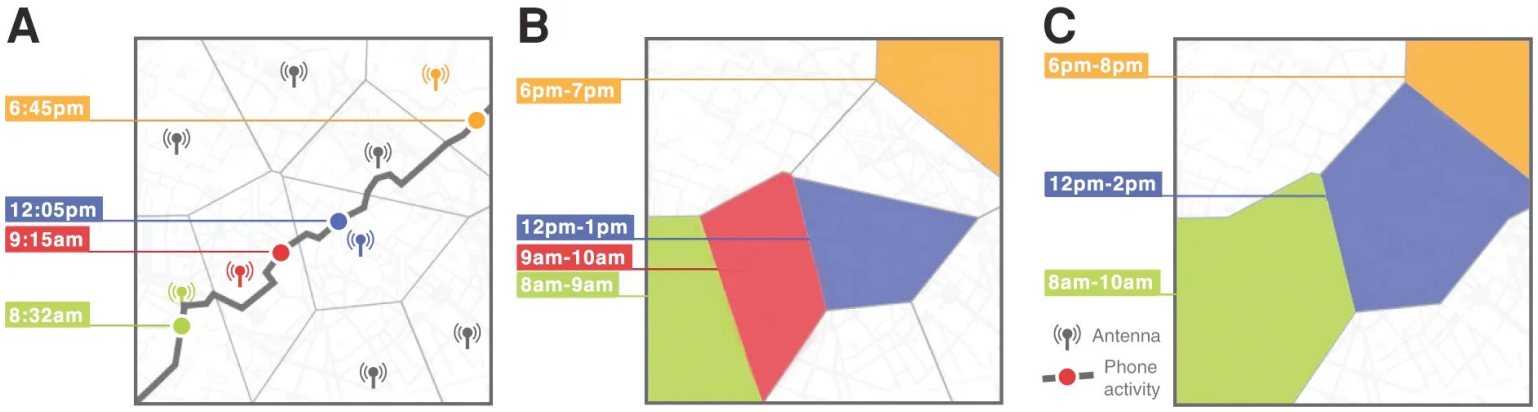
- Mobility traces are unique
- Mobility traces are collected widely:
  - Call data records (CDRs)
    - Each time you make a phone call or send an SMS your telco records the ID of a cell you are associated with
  - GPS traces collected by mobile apps
    - Often without a clear purpose
    - Often shared with (unknown) third parties
  - WiFi/BT associations collected by enterprises

Mobility as Identifier

- "Unique in the Crowd: The privacy bounds of human mobility" De Montjoye et al, 2013
- Dataset:
  - 15,000,000 data points (CDRs)
  - 1,500,000 individuals
  - Coarse resolution:
    - 1 hour time slots
    - Cell ID location resolution
- Conclusion:
  - Four points uniquely identify 95% of the individuals!

Mobility as Identifier

- “Unique in the Crowd: The privacy bounds of human mobility” De Montjoye et al, 2013



- Making the dataset coarser does not help much
- What if there were more users?
  - Cellphone base station density increases with the density of users – no effect on identifiability

Ethics in Mobile Sensing

University *of Ljubljana*
Faculty *of Computer and Information Science*

Ethical Computer Science

- ACM Code of Ethics
  - https://www.acm.org/code-of-ethics
  - General guidelines for responsible research and practice in computer science
- Consult domain experts
  - Do not work on mHealth apps without a medical expert
  - Do not work on social network apps without a sociologist/psychologist

You must be familiar with these guidelines!

Ethical Computer Science

- Comply to legal requirements
  - GDPR regulates personal data collection in EU
    - Disclose any data collection
    - Use only for the given purpose
    - Keep collection to a minimum
    - Right to forget
    - Anonymize and secure
    - Disclose any data breaches
- Comply to institutional requirements
  - Institutional review board (IRB) approval is needed for any personal data collection study (even your MS thesis) at UL FRI

Make Sensing More Ethical

- Sensing modalities
  - Sense only the modalities you need for the app to function properly
  - Sense less identifiable modalities when possible
    - E.g. to infer sleeping: sense time of day and phone-on-charge, instead of sound and location
  - Hierarchical sensing – turn more intrusive sensor when a less intrusive one detects an interesting event

Make Sensing More Ethical

- Sensing accuracy and frequency
  - Use coarser sampling when possible
    - E.g. a weather app does not need fine location
  - Sense when app in the foreground
    - New versions of Android are trying to limit background sensing anyways
  - Sense with the lowest possible frequency/sample size that still provides the neccessary information for your app
    - E.g. noise measurement app could work even with 1s-long recordings

Make Sensing More Ethical

- Minimize data storage
  - Process and discard the data when possible
  - Use internal storage in Android
  - Encrypt local files using Security library
    - `EncryptedFile`
  - Do not expose unintended data to other apps
    - `android:exported=false` for ContentProviders

Make Sensing More Ethical

- Protect data transfer
  - On-device processing whenever possible
    - Use deep learning models optimized for the mobile, instead of sending the data to the cloud
  - Use encrypted connection
    - `HttpsURLConnection`
    - `SSLSocket` that uses TLS under the hood

Ethical Data Processing – Preserving Privacy

- Data anonymization
  - Removing all (obviously) identifiable data, such as locations (use semantic locations instead), names, device IDs, MAC addresses, demographics
- Data anonymization is not enough
  - Alternative authentication mechanisms
    - Gait-based authentication, mobility patterns
  - Data from different datasets can be merged
    - E.g. Netflix challenge failure

Ethical Data Processing – Preserving Privacy

- Third party privacy
  - Users whose data is being collected might not be the same as the users who gave the consent
  - Involved users might not even be aware of their data being sensed
  - Especially prevalent in:
    - Location sensing
    - Wireless sensing
    - Video/voice recording

Ethical Data Processing – Differential Privacy

- Differential privacy:
  - *"Imagine you have two otherwise identical databases, one with your information in it, and one without it. Differential Privacy ensures that the probability that a statistical query will produce a given result is (nearly) the same whether it's conducted on the first or second database."*
    From https://blog.cryptographyengineering.com/2016/06/15/what-is-differential-privacy/
  - Aggregate results with or without a single user's data are the same
  - Apple uses differential privacy for Siri improvements

Ethical Data Processing – Algorithmic Bias

NEWS · 24 OCTOBER 2019 · UPDATE 26 OCTOBER 2019

Millions of black people affected by ra algorithms

Study reveals rampant racism in decision-making software use to correct it.

AI is sending people to jail—and getting it wrong

Using historical data to train risk assessment tools could mean that machines are copying the mistakes of the past.

Recruitment algorithms are 'infected with biases', AI expert warns

By **Ashleigh Webber** on 13 Dec 2019 in **Unconscious bias**, **Artificial Intelligence**, **Equality & diversity**, **Latest News**, **Recruitment & retention**, **Pre-employment screening**

Shutterstock

e-changing" decisions, such as recruitment, should be y are often "infected with biases", according to an ce.