Protecting sensory data against sensitive inferences

Mohammad Malekzadeh, Richard G. Clegg, Andrea Cavallaro, Hamed Haddadi

Published in: *Workshop on Privacy by Design in Distributed Systems (W-P2DS) 2018*

Centre for Intelligent Sensing
Queen Mary University of London
Smart devices measure more and more data every generation.
An example

Google ATAP project Abacus

> 1500 Volunteers

LG Nexus 5

Request

Volunteer Recruiter

Personal Data

Aggregated Data

- **Goal:** using biometric patterns, like motion, instead of password

Ref: DOI 10.1109/ACCESS.2016.2557846
An example

Google ATAP project Abacus

> 1500 Volunteers

LG Nexus 5

Request

Hi Mohammad - unfortunately, no. This dataset contains sensitive user data and cannot be publicly released.

- **Goal:** using biometric patterns, like motion, instead of password
MotionSense Dataset

- Same Activity Set: 6 ADL activities
- Same Place
- Same Phone in the Front Pocket
- Accelerometer and Gyroscope

- 24 Different Subjects:
  - Gender: 14 male - 10 female
  - Age: [18 – 40] years old
  - Weight: [45, 105] kg
  - Height: [160, 195] cm
MotionSense Dataset

![Graph showing Accelerometer (magnitude) Data for Alice and Bob across different activities: Downstairs, Upstairs, Walking, and Jogging.](image)
Visualisation

![t-SNE Component 1](image1)
![t-SNE Component 2](image2)
Autocorrelation
Classification

- 1-D Accelerometer\(_{\text{magnitude}}\): (50Hz)
- Time-Window 5 second
- Deep Convolutional Network

<table>
<thead>
<tr>
<th></th>
<th>Classification Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>activity</td>
<td>~ 98%</td>
</tr>
<tr>
<td>gender</td>
<td>~ 96%</td>
</tr>
<tr>
<td>Identity</td>
<td>~ 89%</td>
</tr>
</tbody>
</table>
Proposed Framework

\[
G^*(.) = \arg\min_{G(.) \in \mathcal{F}} \left( p\left( I_s\left( \hat{S}_d \right) \right) - p\left( I_n\left( \hat{S}_d \right) \right) \right)
\]

Sensitive Inferences

Non-Sensitive Inferences
After Transformation

Alice (Original)

Alice (Anonymised)

Downstairs  Upstairs  Walking  Jogging
Transformed Data

Alice

Bob

Downstairs  Upstairs  Walking  Jogging
## Results

<table>
<thead>
<tr>
<th>Setting</th>
<th>Dataset</th>
<th>Inf.</th>
<th>$S_d$</th>
<th>$\hat{S}_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train: 2 out of 3</td>
<td>MotionSense</td>
<td>activity</td>
<td>95.08</td>
<td>93.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>gender</td>
<td>95.15</td>
<td>49.32</td>
</tr>
<tr>
<td>Train: 3/4 subjects</td>
<td>MotionSense</td>
<td>activity</td>
<td>94.31</td>
<td>90.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>gender</td>
<td>93.74</td>
<td>49.83</td>
</tr>
<tr>
<td></td>
<td>MobiAct</td>
<td>activity</td>
<td>86.33</td>
<td>85.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>gender</td>
<td>75.35</td>
<td>52.16</td>
</tr>
<tr>
<td></td>
<td>MobiAct</td>
<td>activity</td>
<td>70.49</td>
<td>65.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>gender</td>
<td>66.18</td>
<td>45.54</td>
</tr>
</tbody>
</table>
Next Steps

• **Practical:**
  – The **Cost** of the solution on **Edge** devices?
  – Removing identifiable motion patterns.

• **Theoretical:**
  – Provide a **statistical guarantee** (probabilistic bound)
    - Differential Privacy : **Composition Theorem**?
    - Mutual Information : **Joint Distributions**?
Thanks!

Repository of the MotionSense Dataset: bit.ly/eli-dw18

[QR Code]