Future of Privacy Notice and Choice
Panel on:

Leveraging Technology to Improve Notice and Choice:

Panelists:

Norman Sadeh, CMU (moderator)
Joanne Futsch, TRUSTe
Sebastian Zimmeck, Columbia U.
Maritza Johnson, Facebook
Sean Maguire, Microsoft
Context & Objective

☐ “Notice and Choice” in its current form is unworkable

- Length of policies
- Ambiguous language
- Cognitive and behavioral biases
- Does not scale...Internet of Things
- etc.

☐ How could technology help overcome these challenges?
Future of Privacy Notice and Choice

Towards Personalized Privacy Assistants

Norman Sadeh
Professor, School of Computer Science
Director, Mobile Commerce Lab.
Co-Director, MSIT in Privacy Engineering Program
Carnegie Mellon University

www.cs.cmu.edu/~sadeh  ---  sadeh@cs.cmu.edu
Some of Our Own Research

- No one reads privacy notices
- ...and those few people who try are still left wondering what they have read
- Will never work with the Internet of Things
- Instead:
  - Leverage advances in AI, ML, NLP, etc.
  - ...Those very same techniques used to invade our privacy today
How Do People Feel About This?

Percentages of people surprised by an App’s Permission Requests

- **Facebook (rank 1):**
  - Location: 60%
  - Device ID: 0%
  - Contact List: 40%

- **Angry Bird (rank 3):**
  - Location: 80%
  - Device ID: 40%
  - Contact List: 0%

- **Pandora (rank 4):**
  - Location: 80%
  - Device ID: 40%
  - Contact List: 0%

- **Brightest Flashlight (rank 78):**
  - Location: 95%
  - Device ID: 95%
  - Contact List: 0%

Identifying the Purpose of Permissions

- Androguard Reverse Engineering Tool
- Amazon EC2
  - 2035 instance hours $\rightarrow$ 1.23 minutes/app
- 89,903 apps successfully analyzed
  - 83.05% successful rate
  - Failures primarily due to code obfuscation

## Categories of 3rd Party Libraries (Top 400)

<table>
<thead>
<tr>
<th>Targeted Ads</th>
<th>Mobile Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNS</td>
<td>Utility</td>
</tr>
<tr>
<td>Customized UI</td>
<td>Content Host</td>
</tr>
<tr>
<td>Component</td>
<td>Secondary Market</td>
</tr>
<tr>
<td>Game Engine</td>
<td></td>
</tr>
<tr>
<td>Payment</td>
<td></td>
</tr>
</tbody>
</table>
## Permission Breakdown by Purpose

<table>
<thead>
<tr>
<th>Permission</th>
<th>Internal Use</th>
<th>Targeted Ads</th>
<th>Mobile Analytics</th>
<th>SNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERNET</td>
<td>41.33%</td>
<td>47.48%</td>
<td>20.71%</td>
<td>16.30%</td>
</tr>
<tr>
<td>LOCATION</td>
<td>17.48%</td>
<td>72.94%</td>
<td>26.08%</td>
<td>6.07%</td>
</tr>
<tr>
<td>PHONE_STATE</td>
<td>24.55%</td>
<td>74.40%</td>
<td>16.04%</td>
<td>6.35%</td>
</tr>
<tr>
<td>READ_CONTACTS</td>
<td>52.07%</td>
<td>45.76%</td>
<td>-</td>
<td>2.81%</td>
</tr>
<tr>
<td>BLUETOOTH</td>
<td>86.54%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SMS</td>
<td>63.33%</td>
<td>38.81%</td>
<td>-</td>
<td>1.19%</td>
</tr>
<tr>
<td>GET_ACCOUNTS</td>
<td>32.51%</td>
<td>4.95%</td>
<td>-</td>
<td>8.04%</td>
</tr>
<tr>
<td>CAMERA</td>
<td>30.06%</td>
<td>17.45%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RECORD_AUDIO</td>
<td>91.91%</td>
<td>9.51%</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

In other words, **72.94% of apps requiring access to your location, use it for targeted ads.**

Some apps require some permissions for multiple purposes.
Android Permissions: Purpose Matters!

Users’ Average Preferences
White → comfortable
Red → uncomfortable

One Size-Fits-All Defaults Won’t Work

Users’ Average Preferences
White → comfortable
Red → uncomfortable

Variance among Users
Darker yellow → larger variance
Hierarchical Clustering

Canberra distance with average linkage method with $k=4$
Identifying a User’s Privacy Profile

- Asking users a small set of questions

Q1

Q2

Q3

Advanced Users

Conservatives

The unconcerned

Fence-sitters
Pure Prediction vs. Interactive Model

With more labeling of users, we can increase the accuracy of our predictions.

If users can label an additional 10% of their permission decisions, the prediction accuracy will climb from 87.8% to 91.8%...and that’s only 6 questions...

At 20% (about 12 questions), accuracy climbs to 94%!

Vision

- Help scale to interactions with a large number of apps and services
- Learn user preferences
- Can very selectively enter in dialogues with users and nudge them towards safer practice

Did you know that over the past 24 hours your apps shared your location 37 times with 5 different Profiling Companies?
Usable Privacy Policy Project (usableprivacy.org)

- **Semi-automated understanding of privacy policies**
  - Crowd-sourcing + ML + NLP
  - How much can be automated, how to best combine crowd sourcing & technology, task decomposition

- **User privacy preference modeling and nutrition labels**
  - We only need to extract a small number of key policy features
  - Present key features through simple UI (akak nutrition label, icon, letter grade), simplify through personalization & experiment with just-in-time
Future of Privacy Notice and Choice
Panel on:

Leveraging Technology to Improve Notice and Choice:

Panelists:

Norman Sadeh, CMU (moderator)
Joanne Furtsch, TRUSTe
Sebastian Zimmeck, Columbia U.
Maritza Johnson, Facebook
Sean Maguire, Microsoft