Where are your data going?
Feedback Mechanism for Data Leakage in Mobile Devices

Dr Tristan Henderson
Department of Computer Science
University of St Andrews

Onur Mimaroglu

"Facebook" Would Like to Access your Heart Rate

Cancel  OK
The Mobile Era

**For developers**
- Quick and wider access to end users
- Well-defined application programming interfaces (APIs)
- Increased computation power of mobile devices
- Additional features, such as sensors

**For consumers**
- Centralized service software for downloading third-party applications (App Stores)
- Easy access to App Stores
- Downloading apps with little or no prices
- Increased disk space in mobile devices
- Readily available and fast Internet connection

As of May 2015, there are more than 1.5 million apps in Google App Store and 50 billion apps in total had been downloaded in 2013 with both these numbers continue to grow at an increasing rate.
Privacy Issues in Mobile

- Rogue apps put users’ data at risk
- Easy to develop and publish mobile apps
- Permission models of mobile platforms cannot guarantee users’ privacy
- Mobile devices store large amount of privacy-sensitive data
- The data is often opaque to its owner
- Low quality apps
- No control how, when, by whom, the data is used by apps
- No legitimate reasons, no justification why an app needs data

Users need more control over their personal data!
Research Questions

How can we give mobile users more control over their personal data that is subject to collection and misuse by mobile apps?

**R1** How can we restrict dynamic and unpredictable data accesses of mobile applications at run-time without affecting the user-desired functionality of those apps?

**R2** Can we affect users’ privacy decisions at run-time?

**R3** How can we develop an unobtrusive yet salient real-time feedback mechanism?

**R4** How can we convert quantitatively measured privacy exposure into a qualitative notification system?

**R5** Does immediate feedback provide better understanding of possible privacy risks?
The rest of the presentation...

1) Types of Related Work
2) Solution
   – Data Access Detection
   – Feedback Mechanisms
   – Privacy Manager
3) Evaluation
   – Modularity
   – Effectiveness
   – Performance
4) Limitations and Future Work
5) Conclusion
Related Work

1) The systems which detect, monitor or track unwanted data accesses
   – No feedback about how, when and to what extent privacy-sensitive data is exposed to third-parties
   – No privacy manager to give user finer-grained control

2) The studies which aim to raise user awareness and perception through privacy displays
   – Aim is to make user informed, no implementation, often studies

3) The built-in or open source tools that let users to manage privacy settings
   – No access detection, revoke permissions, require custom ROM, no feedback
   – CyanogenMod, Paranoid Android, etc.

No solution combines all three: **Access Detection, Privacy Manager, Feedback Mechanism**
Solution – Android & Xposed
Solution – Overall Architecture
In order to notify users about data accesses, first accesses should be detected.

To develop a generic access detection system, the system hooks the functions of resource managers. 

One should explore available functions in API documents to discover whether any of them might cause personal data leakage.
<table>
<thead>
<tr>
<th>Category</th>
<th>Number of Functions</th>
<th>The Most Sensitive Function</th>
<th>Sensitiveness Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accounts</td>
<td>22</td>
<td>getAccounts</td>
<td>10</td>
</tr>
<tr>
<td>Browser</td>
<td>2</td>
<td>accessAllDownloads</td>
<td>9</td>
</tr>
<tr>
<td>Calendar</td>
<td>4</td>
<td>ContentResolver.query</td>
<td>8</td>
</tr>
<tr>
<td>Calling</td>
<td>16</td>
<td>readCallLog</td>
<td>5</td>
</tr>
<tr>
<td>Clipboard</td>
<td>11</td>
<td>getPrimaryClip</td>
<td>1</td>
</tr>
<tr>
<td>Contacts</td>
<td>8</td>
<td>readContacts</td>
<td>9</td>
</tr>
<tr>
<td>Dictionary</td>
<td>1</td>
<td>readUserDictionary</td>
<td>3</td>
</tr>
<tr>
<td>Email</td>
<td>2</td>
<td>readContentProvider</td>
<td>10</td>
</tr>
<tr>
<td>Identification</td>
<td>21</td>
<td>getDeviceId</td>
<td>6</td>
</tr>
<tr>
<td>Internet</td>
<td>31</td>
<td>getAllAddresses</td>
<td>4</td>
</tr>
<tr>
<td>IPC</td>
<td>1</td>
<td>Binder</td>
<td>1</td>
</tr>
<tr>
<td>Location</td>
<td>51</td>
<td>requestLocationUpdates</td>
<td>9</td>
</tr>
<tr>
<td>Media</td>
<td>20</td>
<td>MediaRecorder.start</td>
<td>9</td>
</tr>
<tr>
<td>Messages</td>
<td>12</td>
<td>readSMS</td>
<td>7</td>
</tr>
<tr>
<td>Network</td>
<td>21</td>
<td>getConnectionInfo</td>
<td>4</td>
</tr>
<tr>
<td>NFC</td>
<td>5</td>
<td>getNfcAdapter</td>
<td>1</td>
</tr>
<tr>
<td>Notifications</td>
<td>3</td>
<td>RECEIVE</td>
<td>1</td>
</tr>
<tr>
<td>Overlay</td>
<td>1</td>
<td>addWindow</td>
<td>1</td>
</tr>
<tr>
<td>Phone</td>
<td>43</td>
<td>getLineNumber</td>
<td>10</td>
</tr>
<tr>
<td>Sensors</td>
<td>16</td>
<td>registerListener</td>
<td>8</td>
</tr>
<tr>
<td>Shell</td>
<td>6</td>
<td>sh command</td>
<td>1</td>
</tr>
<tr>
<td>Storage</td>
<td>12</td>
<td>open / sdcard</td>
<td>10</td>
</tr>
<tr>
<td>System</td>
<td>60</td>
<td>queryUsageStats</td>
<td>7</td>
</tr>
<tr>
<td>View</td>
<td>11</td>
<td>load/postUrl</td>
<td>2</td>
</tr>
<tr>
<td>No Category</td>
<td>63</td>
<td>-</td>
<td>0</td>
</tr>
</tbody>
</table>
Solution – Permission Priorities

Priorities of individual permission categories is given as:

\[ P_i = \frac{w_i}{\sum_{i=1}^{n} w_i} \]

where \( P_i \) is the priority of the \( i^{th} \) category, \( w_i \) is the weight of the \( i^{th} \) category, where \( 1 \leq i \leq n \) and \( n \) is the total number of categories.

\[ \text{total exposure} = \text{coef} \times \left( \sum_{i=1}^{n} t_i \cdot P_i \right) \]

where \( t_i \) is the total number of access attempts of \( i^{th} \) category.
Solution — Feedback Mechanisms

- Intuitive, nonintrusive, incremental, quantifiable yet salient feedback mechanism which can convey exposure level and exposure details

<table>
<thead>
<tr>
<th>Feedback Method</th>
<th>Intuitive</th>
<th>Quantifiable</th>
<th>Intrusive</th>
<th>Salient</th>
<th>Sticky</th>
<th>Incremental</th>
<th>Informative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flash</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>None</td>
</tr>
<tr>
<td>Vibration</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>None</td>
</tr>
<tr>
<td>Audio</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Nudges</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>None</td>
<td>Medium</td>
</tr>
<tr>
<td>Filter</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>None</td>
</tr>
</tbody>
</table>
Evaluation – Modularity & Extensibility
# Evaluation — Effectiveness

<table>
<thead>
<tr>
<th>App</th>
<th>Ident</th>
<th>Sens</th>
<th>Loc</th>
<th>Phon</th>
<th>Netw or Internet</th>
<th>Calen</th>
<th>Stor</th>
<th>Media</th>
<th>Shell</th>
<th>Acnt</th>
<th>NFC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AntiVirus</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>43</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Barcode Scanner</td>
<td>1</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>186</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>BBC News</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>455</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BBC Weather</td>
<td>10</td>
<td>1</td>
<td>6/1</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Blackjack</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cestos</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Coupons</td>
<td>26</td>
<td>3</td>
<td>143/72</td>
<td>3</td>
<td>46</td>
<td>1</td>
<td>221</td>
<td>0</td>
<td>1/1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DasTelefonbuch</td>
<td>14</td>
<td>0</td>
<td>25</td>
<td>4</td>
<td>226</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Evernote</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FlappyPig</td>
<td>18</td>
<td>4</td>
<td>2</td>
<td>14</td>
<td>30</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Glob Traf Radar</td>
<td>36</td>
<td>11</td>
<td>21</td>
<td>6</td>
<td>1670/348</td>
<td>0</td>
<td>52</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Horoscope</td>
<td>7</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>47/1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ixMAT Scanner</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>190</td>
<td>0</td>
</tr>
<tr>
<td>Layar</td>
<td>22</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>4</td>
<td>88</td>
<td>5</td>
</tr>
<tr>
<td>Myspace Music</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Paper Toss 2.0</td>
<td>7</td>
<td>3</td>
<td>23</td>
<td>9</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Showtimes</td>
<td>23/12</td>
<td>0</td>
<td>16/12</td>
<td>0</td>
<td>57/33</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>72/36</td>
</tr>
<tr>
<td>Tasks</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Traffic Jam</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>5</td>
<td>75</td>
<td>0</td>
<td>22</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Wisdom Quotes</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Flashlight</td>
<td>18</td>
<td>4</td>
<td>1</td>
<td>10</td>
<td>71</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

20/07/2015

ONUR MIMAROGLU
Evaluation – Performance – Virtual Machine

Without the System  With the System

CAFFEINEMARK TESTS

CAFFEINEMARK SCORE

SIEVE LOOP LOGIC STRING FLOAT METHOD OVERALL

Without the System: 3854, 24128, 18161, 7949, 10716, 8399, 10320
With the System: 3680, 24091, 17047, 7763, 10608, 7522, 9944
Evaluation – Performance – Startup Durations of Apps

Average overhead to startup durations of 32 apps is 50% at worst-case
Limitations & Future Work

- No justification why an app makes an access attempt
- No data flow e.g., where is exactly the data sent to?
- No user study
- Privacy preferences can be adjusted automatically
  - User’s previous choices and context e.g., time and location
  - Crowdsourced recommendation
- Sending fake data may lead to unexpected behaviours in apps
- Users can benefit from information visualization, e.g., 2-D graphs
Contributions

(1) We develop a mobile system for detecting and restricting unwanted data accesses of Android apps at run-time.

(2) We build a privacy manager to provide users with high granularity in their privacy decisions.

(3) We develop and evaluate various in-situ feedback mechanisms to help users make better informed privacy decisions by conveying exposure level.

(4) We present metrics that embody other factors which in their nature should have an impact on exposure level.
Conclusion

We believe our work will help users to make

✓ More informed
✓ Finer-grained
✓ Runtime

privacy decisions on mobile devices
thanks


