Privacy Implications of Automated GPS Tracking and Profiling

Muhammad Usman Iqbal

Samsung Lim

The University of New South Wales
Presentation Outline ..

- GPS and the Smart Car Revolution
- GPS-based Profiling
- The Profiling Experiment
- Related Privacy Issues
- Conclusion
The Automobile Evolution

• Past
  – Mechanical & hydraulic Parts
  – Unreliable
  – Uncomfortable

• Present
  – Uses computers
  – Telematics and GPS navigation
  – Keyless ignitions, ABS brakes
  – Various sensors

Source: http://patentpendingblogs.com/patent_pending_blog/automobile/index.html

GPS and The Smart Car Revolution

- Cars would be one of the major users of GPS
- Traditionally used for route navigation
- Recent uses include
  - GPS as alibi
  - Covert Surveillance
  - Mobility-based charging

  – And more recently...Congestion Charging
GPS Alibi

- Sets a legal precedent of GPS admissibility in NSW courts
- Lead to charges being dropped and fine overturned

Source: Sydney Morning Herald (12 March 2007)
Covert Surveillance

- Law Enforcement
  - US vs. Garcia
- Employers
- Curious spouse
  - George Ford Case

Motorists have no expectation of privacy on public roads

Source: TrackStick Pro userguide, pp 30
Mobility-based Charging

- Insurance
- Toll Collection
- Road Tax
- Parking

- And Congestion-charging
Data Collection and Retention
Data Collection and Retention

• Sources of GPS data
  – Surveillance
  – Commercial projects (insurance, LBS, etc.)

• Decreasing cost of data storage
  – Data held indefinitely
  – Gives sufficient time to seek alternate use

• Significant privacy threats
  – No information or consent
  – Misreporting
  – Profiling ..
Profiling

“.. a technique whereby a set of characteristics of a particular class of person is inferred from past experience, and dataholdings are then searched for individuals with a close fit to that set of characteristics”

Roger Clarke, 1993.

GPS-based Profiling

Drawing inferences about individuals based on their GPS logs.

- Inferences have been made
  - Driving behaviour and Transport studies
Profiling Paradigm

Adversary

Passive Surveillance

Information Processing

Summary Profiles
Automated GPS Profiling

- 1020 Google Scholar hits for
  - “GPS + Data mining + Privacy”
  - Very few efforts to critically evaluate the privacy threats of profiling

- To demonstrate
  - Wealth of inferences may be drawn
  - Which may not necessarily be correct
  - Violate user’s privacy
  - Automated: future abuses would be automated
The Experiment

• Collect data with a passive GPS device
• Process this data
  – Identify “significant locations”
  – Use data mining techniques, heuristics and rules
  – Correlate information from other databases
  – Infer home, personality, pattern and road behaviour
• Create summary profiles using tables and maps
• Compare with actual by interviewing volunteers
The Tools

• Passive GPS Surveillance device
• Software
  – PSMA’s GeoCoded National Address File
  – PSMA’s transport Maps
  – ESRI ArcGIS and ArcObjects
  – Visual Basic
  – PostGIS
• Heuristics
Surveillance

- Volunteers included
  - Academic staff
  - Administrative staff
  - Research student
  - Undergrad student
- Data collected for at least one week
- Device requires no input from user
- Plugs into the cigarette lighter jack
- Logs data 4-6 times per minute
Information Processing

• Import “anonymous” GPS data
• Cleanse data
• Identify significant locations

<table>
<thead>
<tr>
<th>Rec #</th>
<th>Date</th>
<th>Time</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Altitude</th>
<th>Status</th>
<th>Course</th>
<th>GPS Fix</th>
</tr>
</thead>
<tbody>
<tr>
<td>134</td>
<td>03/17/2007</td>
<td>11:15</td>
<td>-33.9120°</td>
<td>151.1194°</td>
<td>24.7 m</td>
<td>31 kph</td>
<td>NE</td>
<td>Y</td>
</tr>
<tr>
<td>135</td>
<td>03/17/2007</td>
<td>11:15</td>
<td>-33.9116°</td>
<td>151.1199°</td>
<td>25.4 m</td>
<td>31 kph</td>
<td>NE</td>
<td>Y</td>
</tr>
<tr>
<td>135</td>
<td>03/17/2007</td>
<td>11:15</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>Power Off</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>140</td>
<td>03/17/2007</td>
<td>11:22</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>Power On</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>141</td>
<td>03/17/2007</td>
<td>11:22</td>
<td>-33.9054°</td>
<td>151.1275°</td>
<td>0.0 m</td>
<td>0 kph</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

Black dots - significant locations
Information Processing

Vehicle parked – Significant location
Home Determination Algorithm

• Exclude insignificant locations using temporal heuristics to identify home and work locations
  – Morning destinations
  – Afternoon/Evening destinations
• A set of locations short-listed as potential candidates
• Use **PSMA’s GNAF** data address file for NSW
  – Apply a nearness query for a residential address to these points using PostGIS
  – The statistical mode considered as the most likely address
  – Classic example of correlating databases
Home Determination Algorithm

Reported Results

<table>
<thead>
<tr>
<th>Home location</th>
<th>Volunteer C</th>
<th>Volunteer B</th>
<th>Volunteer Y</th>
<th>Volunteer J</th>
<th>Volunteer U</th>
</tr>
</thead>
<tbody>
<tr>
<td>Street number(s) inferred</td>
<td>7</td>
<td>39</td>
<td>24, 25</td>
<td>44</td>
<td>53</td>
</tr>
<tr>
<td>Actual street number</td>
<td>11</td>
<td>39</td>
<td>22</td>
<td>Different street</td>
<td>51</td>
</tr>
</tbody>
</table>

Assumption: 2 address buffer
Inferring Accuracy

Street address for volunteer incorrectly guessed by the home determination algorithm
## Summary Profile

<table>
<thead>
<tr>
<th>Work and commute profile</th>
<th>Volunteer C</th>
<th>Volunteer B</th>
<th>Volunteer Y</th>
<th>Volunteer J</th>
<th>Volunteer U</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total GPS records</td>
<td>5240</td>
<td>1997</td>
<td>2330</td>
<td>4812</td>
<td>2147</td>
</tr>
<tr>
<td>Total Distance</td>
<td>301 km</td>
<td>174.59 km</td>
<td>172 km</td>
<td>284.9 km</td>
<td>149.72 km</td>
</tr>
<tr>
<td>Average distance</td>
<td>27.38 km</td>
<td>34.59 km</td>
<td>31.2 km</td>
<td>40.7 km</td>
<td>37.43 km</td>
</tr>
<tr>
<td>Total travel time</td>
<td>12 hr 45 m</td>
<td>4 hr 25 m</td>
<td>5 hr 1 m</td>
<td>11 hr 44 m</td>
<td>4 hr 51 m</td>
</tr>
<tr>
<td>Average travel time</td>
<td>1 hr 10 m</td>
<td>52 m</td>
<td>54 m</td>
<td>1 hr 40 m</td>
<td>1 hr 12 m</td>
</tr>
<tr>
<td>Max Speed</td>
<td>101 kph</td>
<td>83 kph</td>
<td>86 kph</td>
<td>98 kph</td>
<td>91 kph</td>
</tr>
<tr>
<td>Average Speed</td>
<td>32 kph</td>
<td>45 kph</td>
<td>39 kph</td>
<td>33 kph</td>
<td>39 kph</td>
</tr>
<tr>
<td>Average time leaves home</td>
<td>7:33 am</td>
<td>8:21 am</td>
<td>9:10 am</td>
<td>07:46 am</td>
<td>9:54 am</td>
</tr>
<tr>
<td>Average time leaves work</td>
<td>3:30 pm</td>
<td>5:09 pm</td>
<td>4:54 pm</td>
<td>08:58 pm</td>
<td>5:07 pm</td>
</tr>
<tr>
<td>Average time arrives at work</td>
<td>8:03 am</td>
<td>8:55 am</td>
<td>9:32 am</td>
<td>08:40 am</td>
<td>10:15 am</td>
</tr>
<tr>
<td>Average time at work</td>
<td>7 hr 58 min</td>
<td>8 hr 10 min</td>
<td>7 hr 25 min</td>
<td>12 hr 18 m</td>
<td>6 hr</td>
</tr>
<tr>
<td>Parks car in</td>
<td>University parking lot</td>
<td>University parking lot</td>
<td>University parking lot</td>
<td>University parking lot</td>
<td>Around university</td>
</tr>
<tr>
<td>Type of person</td>
<td>Academic or Support</td>
<td>Academic or Support</td>
<td>Academic or Support</td>
<td>Research Student</td>
<td>Undergrad Student</td>
</tr>
</tbody>
</table>
Speed and Acceleration Monitoring

Red spot indicates deceleration of 3 m/s²---- (hard braking)
Why is Profiling Privacy Invasive?

- Experiment demonstrates risks of profiling
- Secondary use of data
- No consent sought
- Can embarrass or cause harm to a person
- Possibility of Mis-Profiling
- Profiling and stereotyping
  - Risky driver for insurance
Concluding Remarks

• Data gathered only for 1 week
  – Insufficient for highly accurate profile generation
• Predicted the personnel type and residence
• Profile may be circumstantial
  – An academic preparing lecture notes
  – Being on campus doesn’t necessarily mean working
  – High speed, brakings due to emergency
• Mobility-pricing and complete GPS log disclosure raises these threats
  – Important to negotiate the use and retention
  – Use aggregated data reporting
Thank You for your time!

- Questions?