Cybercasing the Joint: On the Privacy Implications of Geo-Tagging

Gerald Friedland, Robin Sommer
International Computer Science Institute
Berkeley, CA
fractor,robin@icsi.berkeley.edu
What is Geotagging?

Why Geo-Tagging?

Allows easier clustering of photo and video series as well as additional services.
Why Geo-Tagging?

Part of location-based service hype:
Support for Geo-Tags

Social media portals provide APIs to connect geo-tags with metadata, accounts, and web content.

<table>
<thead>
<tr>
<th>Portal</th>
<th>%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>YouTube</td>
<td>3.0</td>
<td>3M</td>
</tr>
<tr>
<td>(estimate)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flickr</td>
<td>4.5</td>
<td>180M</td>
</tr>
</tbody>
</table>

Allows easy search, retrieval, and ad placement.
Problems

People are unaware of

1. geo-tagging
2. resulting inference possibilities:
   a. high resolution of sensors
   b. large amount of geo-tagged data
   c. easy-to-use APIs allow fast retrieval
Related Work

“Be careful when using social location sharing services, such as FourSquare.”
Mayhemic Labs, June 2010: “Are you aware that Tweets are geo-tagged?”
Can you do real harm?

• Cybercasing: Using online (location-based) data and services to mount real-world attacks.

• Three Case Studies:
Case Study 1: Twitter

- Pictures in Tweets can be geo-located
- From an undisclosed celebrity we found:
  - Home location (several pics)
  - Where the kids go to school
  - The place where he/she walks the dog
  - “Secret” office
- Systematic search: picfog.com
Celebs unaware of Geo-Tagging

Source: ABC News
Celebs unaware of Geotagging

EXIF IFD1

- Compression {0x0103} = JPEG compression (6)
- X-Resolution {0x011A} = 4718592/65536 ===> 72
- Y-Resolution {0x011B} = 4718592/65536 ===> 72
- X/Y-Resolution Unit {0x0128} = inch (2)
- Y/Cb/Cr Positioning (Subsampling) {0x0213} = centered / center of pixel array (1)
- Embedded thumbnail image:

EXIF GPS IFD

- GPS Version ID {0x00} = 0x02,0x02,0x00,0x00
- GPS Latitude Reference {0x01} = N
- GPS Latitude {0x02} = 34/1,12/1,3/1 [degrees, minutes, seconds] ===> 34° 12' 3'' == 34.200833°
- GPS Longitude Reference {0x03} = W
- GPS Longitude {0x04} = [degrees, minutes, seconds] ===> [redacted]° [redacted]' [redacted]''
Google Maps shows Address...
Case Study 2: Craigslist

- Many ads with geo-location otherwise anonymized
- Sometimes selling high-valued goods, e.g. cars, diamonds
- Sometimes “call Sunday after 6pm”
- Multiple photos allow interpolation of coordinates for higher accuracy
Craigslist: Real Example
Geo-Tagging Resolution

iPhone 3G picture

Google Street View

Measured accuracy: +/- 1m
People are Unaware of Geo–Tagging

“For Sale” section of Bay Area Craigslist.com: 4 days: 68729 pictures total, 1.3% geo–tagged

<table>
<thead>
<tr>
<th>#</th>
<th>Model</th>
<th>#</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>414</td>
<td>iPhone 3G</td>
<td>6</td>
<td>Canon PowerShot SD780</td>
</tr>
<tr>
<td>287</td>
<td>iPhone 3GS</td>
<td>3</td>
<td>MB200</td>
</tr>
<tr>
<td>98</td>
<td>iPhone</td>
<td>2</td>
<td>LG LOTUS</td>
</tr>
<tr>
<td>32</td>
<td>Droid</td>
<td>2</td>
<td>HERO200</td>
</tr>
<tr>
<td>26</td>
<td>SGH-T929</td>
<td>2</td>
<td>BlackBerry 9530</td>
</tr>
<tr>
<td>20</td>
<td>Nexus One</td>
<td>1</td>
<td>RAPH800</td>
</tr>
<tr>
<td>9</td>
<td>SPH-M900</td>
<td>1</td>
<td>N96</td>
</tr>
<tr>
<td>9</td>
<td>RDC-i700</td>
<td>1</td>
<td>DMC-ZS7</td>
</tr>
<tr>
<td>6</td>
<td>T-Mobile G1</td>
<td>1</td>
<td>BlackBerry 9630</td>
</tr>
</tbody>
</table>
Case Study 3: YouTube

- Once data is published, the Internet keeps it (in potentially many copies).
- APIs are easy to use and allow quick retrieval of large amounts of data.
- Even simple inference algorithms (across different websites) allow for cybercasing.

Can we find people on vacation in YouTube?
Cybercasing on YouTube

Experiment: Cybercasing using the YouTube API (240 lines in Python)

- Location
- Radius
- Keywords

Users?

Time-Frame
Distance

Query

Results

Filter

Cybercasing Candidates

YouTube
Cybercasing on YouTube

Input parameters

Location: 37.869885, -122.270539
Radius: 100km
Keywords: kids
Distance: 1000km
Time-frame: this_week
Cybercasing on YouTube

Output

Initial videos: 1000 (max_res)

➡ User hull: ~50k videos

➡ Vacation hits: 106

➡ Cybercasing targets: >12
Cybercasing on YouTube

First Day of [redacted] Vacation

1 videos  Subscribe

Overview:

- Initial videos: 1000 (max_res)
- User hull: ~50k videos
- Vacation hits: 106
- Cybercasing targets: >12
Solutions?
Solutions?

- Better Education
- More secure default values
- Blurring
- Scrubbing
- Privacy-preserving APIs and policies
Proposal: Opt−In with Choice of Accuracy

Mockup of a privacy−improved iPhone dialog
Conclusion

● Geo-location offers great opportunities and we should continue to explore them

● However it can pose real-world risks

● Therefore, we should:
  ● Raise the awareness on privacy issues
  ● Discuss policies and interfaces
Questions?

- Are you concerned?

- What is a good trade-off between privacy and utility?

- How can we design policies and APIs to implement the trade-off?