Protecting Location Privacy: It’s Not Who You Know, It’s Where You Go

Michael Solomon, Ph.D.
June 2017
Michael Solomon, Ph.D.

- CISSP PMP CISA
- Solomon Consulting Inc.
  - OpenEdge, Roundtable, Security architecture
    - Since 1988 (Progress Version 4)
  - CyberSecurity Simulation attack team leader
    - Penetration testing, attack detection and response
- Emory University
  - Assured Information Management and Sharing (AIMS)
  - Private location proximity detection research
- University of the Cumberlands
  - Associate Professor, Master of Science in Information Systems Security program
Agenda

- Value – it’s all location, location, location
  - Where are my users?
  - Where have they been?

- Creeping vs Spatial determination
  - Location data analysis
  - Building trajectories

- Protecting user privacy
  - It is a choice
With mobile applications, the *where* matters.
Many devices sense location

- GPS
- Cellular ID
- Wi-fi proximity
- Inertial sensors
- Barometer
- Bluetooth beacons

Don’t forget about the subway!

Primary uses for location

- Gamification
- Social interaction
- Utilitarian
Directions
Start address: 100 NW Couch St, Portland, OR 97209
End address: Hillsboro, OR

Google Transit Beta
100 nw couch st. portland to hillsboro. oregon by 8pm
get directions

Transit Trip Planner
Directions: Drive There - Take Public Transit

Start address:
100 NW Couch St
Portland, OR 97209

End address:
Hillsboro, OR

When: [edit]
Arrivals before 8:00 pm:
7:05 pm - 7:57 pm (51 mins)
7:21 pm - 7:44 pm (54 mins)
7:31 pm - 7:30 pm (55 mins)
7:21 pm - 7:17 pm (55 mins)

Duration:
51 mins in transit
13 mins walking to/from your route

Cost:
$1.80 (vs. $8.41 driving) details

Begin by walking
1. Start at 100 NW Couch St
2. Go to Skidmore Fountain MAX Station (takes about 1 min)

Take the MAX Blue Line (Direction: Hillsboro)
3. 7:05 pm: leave from Skidmore Fountain MAX Station
4. 7:57 pm: arrive at Tuality Hospital/SE 8th Ave MAX Station

End by walking
4. Go to Hillsboro, OR (takes about 12 mins)

These directions are for planning purposes only. You may find that construction, traffic, or other events may alter the actual travel time.
Where are my users?

- Finding a device’s location is easy
  - Assuming the user allows it
- Many frameworks provide methods to return physical location
- HTML5 geolocation object
  - Returned by Navigator.geolocation
  - Handy methods
    - getCurrentPosition()
    - watchPosition()
    - clearWatch()
- But, what about Kendo UI?
Kendo UI - retrieve location coordinates

navigator.geolocation.getCurrentPosition(onSuccess, onError);

// onSuccess Callback
// This method accepts a Position object, which contains the current GPS coordinates
//
var onSuccess = function(position) {
    alert('Latitude: ' + position.coords.latitude + '
    Longitude: ' + position.coords.longitude + '
    Altitude: ' + position.coords.altitude + '
    Accuracy: ' + position.coords.accuracy + '
    Altitude Accuracy: ' + position.coords.altitudeAccuracy + '
    Heading: ' + position.coords.heading + '
    Speed: ' + position.coords.speed + '
    Timestamp: ' + position.timestamp + '');
}

Location data can hold many secrets.
Storing spatiotemporal data

- Fairly easy
- Just more features

Visualizing can be difficult

- Many creative techniques
- Important to prioritize features and dependencies
Trajectories - The real value of spatiotemporal data
Privacy. It matters.
What about privacy?

Confidentiality is about the data

- Access to data
- Intention is to keep data secret
- Allow access only to authorized users

Privacy is about the individual

- Access to the person (or organization)
- Appropriate use of information
  - More than just access to data
- Being free from public attention
- Ability to be left alone

Location data can identify individuals

- Current and past
- What about predictive analytics?
Problems with location privacy

Outright disclosure
- Who knows where you are?
- Who knows where you’ve been?
- You have to divulge your location to consume services
  - Right?

Inference
- Combining partial information to get an answer
  - Checkins
  - Pictures
  - Reviews
  - Past and future
Trajectory analysis quiz

- Could be called “creeper quiz”
- What does the following trajectory imply?
  - Residence -> Elementary school -> Retail Dress shop -> Elementary school -> residence
- What about this trajectory?
  - Residence -> High school -> baseball field -> coffee shop -> residence
<table>
<thead>
<tr>
<th>What can application providers do?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Omit location data</strong></td>
</tr>
<tr>
<td>• Loss of utility and value</td>
</tr>
<tr>
<td><strong>Establish user trust</strong></td>
</tr>
<tr>
<td>• Strong privacy policy</td>
</tr>
<tr>
<td>• Strong controls</td>
</tr>
<tr>
<td>• Incentive for users to trust</td>
</tr>
<tr>
<td>• Service value</td>
</tr>
<tr>
<td><strong>User location privacy</strong></td>
</tr>
<tr>
<td>• Users consume services without divulging locations</td>
</tr>
<tr>
<td>• Multiple approaches (next slides)</td>
</tr>
<tr>
<td><strong>Mutual location privacy</strong></td>
</tr>
<tr>
<td>• Server and user locations are private</td>
</tr>
<tr>
<td>• Sounds odd, but some applications benefit</td>
</tr>
<tr>
<td>• Multiple approaches (next slides)</td>
</tr>
</tbody>
</table>
One class of location processing – proximity detection

- Simple – determining when a device is near some location
  - Area of Interest (AOI)
  - Defined by Data Provider (DP)
  - Can be of any size

- AOIs can be
  - Approach – interesting area
  - Avoid – dangerous area
Private Proximity Detection

Many solutions to keeping user locations private

- While still providing location-based services

Four leading strategies

- Multiple research efforts

<table>
<thead>
<tr>
<th>Location perturbation and Transformation</th>
<th>Access control</th>
<th>Private Information Retrieval (PIR)</th>
<th>Encryption</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Loss of precision</td>
<td>- Limited granularity</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Requires trusted third party</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Server location data is not private</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Computational overhead</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Mutually Private Proximity Detection (MPPD)

Am I near a dangerous area?

Where is the fire?

Are there any short lines nearby?

Are you a premium subscriber?
AOI definitions

AOIs “red”, “blue”, and “green”

AOIs available only to subscribed (paid) users.
Purple represents “red” and “blue” overlap.

Access policy: AOI “red”
“(subscriber=paid) AND (alertType=warn)”

Access policy: AOI “blue”
“(subscriber=paid) AND (alertType=notify)”

Access policy: AOI “green”
“(subscriber=paid) AND (alertType=approach)”

AOI “yellow”

AOIs for free users are more generic (i.e. provide less specific information)

Access policy: AOI “yellow”
“(subscriber=free) AND (alertType=notify)”
MPPD promising approaches

**Bloom Filter**
- “Location privacy without mutual trust: The spatial bloom filter”
- Spatial Bloom filter / Paillier cryptosystem

**Hilbert curve**
- “Hilbert curve-based cryptographic transformation scheme for spatial query processing on outsourced private data”
- Hilbert Aggregation Index (HAI) / Range and kNN queries

**Homomorphic encryption**
- “Secure k-nearest neighbor query over encrypted data in outsourced environments”
- Paillier cryptosystem / Encrypted DB, query
Spatial Bloom Filter (SBF)

- **Bloom filter**
  - Spatial bloom filter / Paillier cryptosystem
  - SBF constructed over multiple sets (AOIs)

- **Overview**
  - Data provider (DP) creates and encrypts SBF (AOIs)
  - User creates and encrypts SBF (location)
  - Service provider (SP) calculates SBF product
  - DO decrypts scrambled result and counts non-zeros
    - DO determines AOI proximity from result, informs user
Hilbert Curve Transformation (MCT)

- Hilbert curve
  - Hilbert Aggregation Index (HAI) (range of cells)
  - Transformed Data Index (TDI) (AOI cells)
  - Range and kNN queries

- Overview
  - DP encodes AOIs in groups of F (fan-out)
  - DP creates index of start/stop cells, then encrypts entries (AES)
  - User (has AES key) requests and decrypts HAI, then requests overlapping TDI entries
  - User filters results to determine proximity
Secure k-Nearest Neighbor (SkNN)

- Homomorphic Encryption
  - Paillier cryptosystem / Encrypted DB, query
- Original protocol not location specific
  - We extended SkNN distance calculation to consider distinct locations
    - Instead of distance between 2 attribute vectors
- DP encrypts AOIs, user encrypts location
- DP/SP use SSED to determine user/AOI proximity
Performance results

(a) Impact of number of AOIs
(b) Impact of AOI size
(c) Impact of grid size
Factors affecting performance

<table>
<thead>
<tr>
<th>Method</th>
<th># AOI</th>
<th>AOI Size</th>
<th>Grid size</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBF</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>SkNN</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>HCT</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
## Privacy guarantees

<table>
<thead>
<tr>
<th>Method</th>
<th>User</th>
<th>Data Provider</th>
<th>Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBF</td>
<td>k-anonymity based on filter size.</td>
<td>User only knows if loc overlaps AOI</td>
<td>DP only sees obfuscated results / cannot correlate to user</td>
</tr>
<tr>
<td>SkNN</td>
<td>DP only learns when user overlaps AOI</td>
<td>User only knows if loc overlaps AOI</td>
<td>DP and SP unable to correlate query to user</td>
</tr>
<tr>
<td>HCT</td>
<td>User location is never shared</td>
<td>k-anonymity based on fan-out (F) value</td>
<td>Limited, SP learns user/AOI proximity with 1/F accuracy</td>
</tr>
</tbody>
</table>
Location privacy. You can do it.