Privacy in Location Aware Systems

http://www.cs.unibo.it/difelice/cas/

Context Aware Systems

Prof. Marco Di Felice
Department of Computer Science and Engineering
University of Bologna
Privacy & Location-awareness

✧ Privacy is a serious issue, but also a trade-off

- Context-aware applications require context acquisition ... i.e. gathering of user-related data (his/her locations, activities, emotions, etc).
- However, without data-sharing, some applications (even those beneficial for the users) cannot be enabled...
Privacy & Location-awareness

✧ Privacy concerns depend from the users’ perceptions

A measure of privacy behavior
Relative admission rates in an experiment testing the impact of different survey interfaces on willingness to answer questions about various sensitive behaviors

Disclosure behavior in online social media
Percentage of profiles publicly revealing information over time (2005-2011)

Alessandro Acquisti Laura Brandimarte, George Loewenstein, Privacy and human behavior in the age of information, Science 30, 2015, https://science.sciencemag.org/content/347/6221/509
Privacy & Location-awareness

✧ Privacy concerns depend from the users’ perceptions

☐ How much is privacy worth?

---

George Danezis, Stephen Lewis and Ross Anderson, How much is privacy worth?, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.61.23
Privacy & Location-awareness

- Definitions
- Privacy Threats
- Privacy Architectures
- Privacy models
- Privacy techniques
- Open issues

Source: https://www.techlicious.com/
Privacy & Location-awareness

“The right to be left alone”
(Luis Brandeis, 1890)

“Privacy is the claim of individuals, groups or institutions to determine for themselves when, how and to what extent information about them is communicated to others…”
(A. F. Westin, 1967)

LOCATION PRIVACY

“A special type of information privacy which concerns the claim of individuals to determine for themselves when, how and to what extent location information about them is communicated to others…”
(M. Duckham, 2006)
Privacy & Location-awareness

LOCATION PRIVACY BASED ON USE-CASE

- **Real-time** use-case
  Users query a Location-based service and expect an immediate answer.
  - Private query over public data (i.e. user location is private, objects of interests are public)
  - Private query over private data (both query location and objects of interests are private)

- **Offline** use-case
  Users send their data to a Location-based service, that is going to create and publish a shared dataset for commercial or non-commercial profits.

- **Batch** use-case
  Users send their data to a Location-based service, that is going to publish an aggregated dataset.
Privacy & Location-awareness

LOCATION PRIVACY BASED ON THE USER REQUIREMENTS

- **User Location Privacy**
  Users want to hide their location information and their query information.

- **User Query Privacy**
  Users do not mind or obligated to reveal their locations, however, users want to hide their queries.

- **Trajectory Privacy**
  Users do not mind to reveal few locations, however, they want to avoid linking these locations together to form a trajectory.
Privacy & Location-awareness

- Definitions
- Privacy Threats
- Privacy Architectures
- Privacy models
- Privacy techniques
- Open issues

Source: https://www.techlicious.com/

Privacy & Location-awareness

Please Rob Me: Site Tells The World When You’re Not Home

Social networking services from Twitter and Foursquare to Yelp and Buzz encourage users to log in and share their location.

It's packaged as a fun way to find your friends and stay social—but what's the cost?

https://www.huffpost.com/entry/please-rob-me-site-tells_n_465966
Privacy & Location-awareness

Strava heat map exposes secret military locations, sparks security fears

The fitness tracker app has inadvertently given out locations and activities of soldiers at military bases throughout the world.

By Ishvina Singh - January 30, 2018

In June 2014, Strava, which offers cycling and running tracking apps for multiple fitness platforms, claimed to be tracking over 1 million activities per week across the globe. In the next couple of years, the GPS tracking startup managed to capture 1 billion activities, 3 trillion lat-long points, and 5% of all land on Earth covered by tiles. Strava naturally wanted to showcase the reach of its huge user community. And that it did by populating its interactive global heat map with a total recorded activity duration of 300,000 years. A

Privacy & Location-awareness

Privacy of location tracking device owners threatened by ‘Trackmageddon’ flaws

Bradley Barth
Rene Millman

A slew of online services used to manage GPS- and GSM-based location tracking devices have been found vulnerable to flaws that could allow attackers to

Apps

Tinder dating app was sharing more of users’ location data than they realised

Security flaw discovered by white-hat hackers in October 2013 was patched by the end of the year

Stuart Dredge
@stuartdredge
Thu 20 Feb 2014
13.14 GMT

It’s a Match!

You and Sarah have liked each other.

Send a Message
Privacy & Location-awareness

Use Google Maps in Incognito mode

You now have more ways to control your privacy on Google Maps. Use Incognito mode when you don't want your activity—like the places you search for or navigate to—to be saved to your Google Account.

Important: When Incognito mode is on, Maps on that device will not:
- Save your browsing or search history in your account, or send notifications.
- Update your Location History or shared location, if any.
- Use your activity to personalize Maps.

Turning on Incognito mode in Maps does not affect how your activity is used or saved by internet providers, other apps, voice search, and other Google services.

Android  iPhone & iPad

Turn on Incognito mode for Google Maps

1. On your Android phone or tablet, open the Google Maps app 🗺️.
2. In the top right, tap your profile picture.
3. Tap Turn on Incognito mode.

Human mobility traces are highly unique, i.e. they are identifiable with minimal prior knowledge!

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of locations</td>
<td>~ 1,700</td>
</tr>
<tr>
<td>Number of timestamps</td>
<td>672</td>
</tr>
<tr>
<td>Number of trajectories</td>
<td>633,798</td>
</tr>
<tr>
<td>Records in the original data</td>
<td>~ 56,000,000</td>
</tr>
<tr>
<td>Records in the cleaned data</td>
<td>~ 43,000,000</td>
</tr>
<tr>
<td>Users in the original data</td>
<td>~ 1,100,000</td>
</tr>
<tr>
<td>Users in the cleaned data</td>
<td>~ 630,000</td>
</tr>
<tr>
<td>Minimum number of points in one trajectory</td>
<td>3</td>
</tr>
<tr>
<td>Maximum number of points in one trajectory</td>
<td>672</td>
</tr>
<tr>
<td>Average number of points in one trajectory</td>
<td>56</td>
</tr>
</tbody>
</table>
Privacy & Location-awareness

1. Infer **users’ Point of Interests (POIs)**

**POIs** → generic locations where users spend some time

- Mobility traces of taxi drivers in the San Francisco Bay Area.
- Consider stopping points of the taxi drivers, and aggregate them via clustering techniques.
- Infer **POIs** of drivers.
- Locate **drivers’ home** and confirm it by satellite views of the area (Google Traffic View and Google Maps).
2. Infer semantic of POIs and of mobility behaviors

Understand what POIs represent

- Mobility traces of users, gathered through the GPS sensor embedded in mobile phones.
- Clustering + Markov Chain analysis + geocoding techniques.
- Infer names of the Markov C. states.
Privacy & Location-awareness

3. Discover social relationships

Compare traces and infer relationships among users

- The attacker is able to eavesdrop the Wi-Fi traffic, getting the RSSI and MAC address of each packet.
- **GOAL 1**: identify an encounter between any two users
- **GOAL 2**: Classify each encounter (friends, classmates and others)
Privacy & Location-awareness

4. Future mobility prediction

Predict places where users are likely to go

- 35 million check-ins made by about 1 million Foursquare users in over 5 million venues across the globe.
- **Predict the next venue a mobile user will visit through a supervised learning approach.**
- Features extracted:
  - transitions between types of places
  - mobility flows between venues
  - spatio-temporal characteristics of user check-ins
Privacy & Location-awareness

5. Allow Users’ Re-identification

Identify physical users from anonymized datasets

- Use the mobility traces of taxi drivers in NY
- Combine pictures of celebrities with the mobility information in order to infer the destination addresses!
Privacy & Location-awareness

- Definitions
- Privacy Threats
- Privacy Architectures
- Privacy models
- Privacy techniques
- Open issues

Source: https://www.techlicious.com/
Privacy & Location-awareness

Location Privacy Protection Mechanisms (LPPM)

Local architecture
The end-device fully implements the LPPM before sending data to the Location-Based Service (LBS).
Privacy & Location-awareness

Location Privacy Protection Mechanisms (LPPM)

✧ **Proxy-based** architecture

LPPM is implemented by a trusted third-party proxy server. The truster server sends data to the Location-Based Service.

---

**Privacy in Location-aware Systems**

**Marco Di Felice**
Privacy & Location-awareness

Location Privacy Protection Mechanisms (LPPM)

✧ Peer-to-Peer (P2P) architecture

Users are engaged in a collaborative privacy protocol before sending data to the Location-Based Service (LBS).

GET REAL GPS LOCATIONS

END-DEVICE(S)

CLUSTER-HEAD

SEND OBfuscated GPS POSITIONS

< f(Lat, Long) >

LOCATION-BASED SERVICE

SPATIAL QUERY PROCESSOR

PRIVACY IN LOCATION-AWARE SYSTEMS

MARCO DI FELICE
Privacy & Location-awareness

- Definitions
- Privacy Threats
- Privacy Architectures
- Privacy models
- Privacy techniques
- Open issues

Source: https://www.techlicious.com/
Privacy & Location-awareness

- Location **cloaking**
  - The user location is blurred into a **region** including the real position.
  - An adversary only knows the **cloaked** region.

Privacy & Location-awareness

K-Anonimity

A user must be indistinguishable among other K-1 users
For an attacker, the probability of re-identification (without any external knowledge) is 1/K

<table>
<thead>
<tr>
<th>Name</th>
<th>Birth</th>
<th>Sex</th>
<th>Zip</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>1970</td>
<td>M</td>
<td>0247</td>
<td>Chest pain</td>
</tr>
<tr>
<td>*</td>
<td>1970</td>
<td>M</td>
<td>0247</td>
<td>Asthma</td>
</tr>
<tr>
<td>*</td>
<td>1972</td>
<td>F</td>
<td>0247</td>
<td>Migraine</td>
</tr>
<tr>
<td>*</td>
<td>1972</td>
<td>F</td>
<td>0247</td>
<td>Asthma</td>
</tr>
</tbody>
</table>
Privacy & Location-awareness

- **K-Anonymity Spatial Cloaking**
  - The cloaked region contains at least $k$ users.
  - The region size must be adaptive (i.e. it depends from the use-case), however the user can specify min-max thresholds.

**Quadtree Spatial Cloaking**

- Divide the space into regions, until a region contains less than $K$ users.
Privacy & Location-awareness

- **Differential Privacy**
  - Formal mathematical framework for guaranteeing privacy protection when analyzing or releasing statistical data.
  - It allows to capture the trade-off between privacy loss and accuracy of the data processing results.
  - **Plausible deniability**: the presence of a specific user in the dataset cannot be ascertained.
  - **Prevent targeted attacks**: reduce the information leaked even in the presence of side knowledge.

Privacy & Location-awareness

**Differential Privacy**

- Several public datasets containing sensitive data are anonymized by removing explicit identifiers (such as name, address and other personally identifiable information) and reducing the geographic information or by using pseudonyms … Is it enough?

Re-identification by linkage attack

People re-identification from environmental health dataset

L. Sweeney, Re-identification Risks in HIPAA Safe Harbor Data: A study of data from one environmental health study, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6344041/
Privacy & Location-awareness

Differential Privacy

INTUITION: Add some (controlled) random noise to the data


Privacy & Location-awareness

- **Differential Privacy**

  - **INTUITION**: Add some *(controlled)* random noise to the data

  - Perform a survey with the following question:
    
    “Have you ever committed a crime? Y/N”

  And then build the dataset by applying this randomization process for each user:

  ![Diagram of the randomization process](https://blogs.oracle.com/datascience/differential-privacy)
Privacy & Location-awareness

☐ **Differential** Privacy

✧ **INTUITION:** Add some *(controlled)* random noise to the data

☐ Perform a survey with the following question:

"Have you ever committed a crime? Y/N"

And then build the dataset by applying this randomization process for each user:

- **Q1.** Can we decide whether user X has committed a crime?
- **Q2.** Can we estimate the real number of users having committed a crime?
- **Q3.** Can we estimate the proportion of users that committed a crime?
Privacy & Location-awareness

- **Differential Privacy**

  ✷ **INTUITION**: Add some *(controlled)* random noise to the data

  “Any information-related risk to a person should not change significantly as a result of that person’s information being included, or not, in the analysis”.

  ![Diagram showing differential privacy](https://blogs.oracle.com/datascience/differential-privacy)
Privacy & Location-awareness

Differential Privacy

INTUITION: Add some (controlled) random noise to the data

A (randomized) algorithm $M: X^n \rightarrow T$ satisfies $\varepsilon$-differential privacy if

$\forall x, x' \in X^n$ that differ on one entry,

$\forall S$ subset of the outcome space $T$,

$$\Pr_M[M(x) \in S] \leq e^\varepsilon \Pr_M[M(x') \in S]$$
Privacy & Location-awareness

- Definitions
- Privacy Threats
- Privacy Architectures
- Privacy models
- Privacy techniques
- Open issues

Source: https://www.techlicious.com/

Privacy & Location-awareness

- **Dummy updates**
  - The user sends $m$ location updates; $m-1$ are fake positions.
  - The user receives $m$ replies from the Location-based Service, and it keeps only the reply related to the true position.

  - **Random** policy [1].
    The next position is decided in a neighbourhood of the current position of the dummy, by adding some random noise.

  - **Collaborative** policy [1].
    The next position is decided randomly in a neighbourhood including the previous dummy and by taking into account the positions of the other peers.
Privacy & Location-awareness

- **K-anonymity & Spatial Cloaking**
  - Make an individual indistinguishable from a group of users.
  - This is achieved by reporting a cloak area, large enough to contain the group necessary to meet the privacy constraints.

- Mobile nodes
- Anonymity Servers (trusted)
- Location-based Servers
Privacy & Location-awareness

- **K-anonimity & Spatial Cloaking**
  - Make an individual indistinguishable from a group of users.
  - This is achieved by reporting a cloak area, large enough to contain the group necessary to meet the privacy constraints.

- **Privacy vs Quality of Service**
- Each mobile node specifies:
  - Anonymity level (k parameter)
  - Spatial tolerance (max cloaking box)
  - Temporal tolerance

---

Privacy & Location-awareness

K-anonymity & Spatial Cloaking
- Make an individual indistinguishable from a group of users.
- This is achieved by reporting a cloak area, large enough to contain the group necessary to meet the privacy constraints.

Upon receiving a message from the mobile node, the anonymous server:
- Decrypts the message
- Removes the identifier
- Perturbs location information through spatial cloaking
Privacy & Location-awareness

- **K-anonimity & Spatial Cloaking**
  - Make an individual indistinguishable from a group of users.
  - This is achieved by reporting a cloak area, large enough to contain the group necessary to meet the privacy constraints.

**CLIQUECLOAK**

Build a graph structure:
- The set of vertexes is the set of nodes
- There exists an arc between node i and node j whether their maximum cloaking box include both of them


Privacy & Location-awareness

K-anonymity & Spatial Cloaking

- Make an individual indistinguishable from a group of users.
- This is achieved by reporting a cloak area, large enough to contain the group necessary to meet the privacy constraints.

Whenever adding a new node m:
- Find a set of nodes that are neighbours of m and have a level of anonymity $\leq k.m$
- Build the cloaking region as the minimum bounding box including m and its neighbours.

source: [2]

Privacy & Location-awareness

K-anonymity & Spatial Cloaking

Make an individual indistinguishable from a group of users.

This is achieved by reporting a cloak area, large enough to contain the group necessary to meet the privacy constraints.

In case a node does not have suitable neighbours

Its messages are buffered by the anonymous server (and not transmitted).

The message is transmitted anyway after a maximum amount of time (temporal tolerance)

source: [2]
Privacy & Location-awareness

- Pseudonym-based mechanisms
  - Each user is using **pseudonyms** in order to anonymize the requests sent to the Location Based Server.

  - Mobile nodes
  - Anonymity Servers (trusted) (storing a list of pseudonyms)
  - Location-based Servers
Privacy & Location-awareness

- **Pseudonym-based mechanisms**
  - Each user is using **pseudonyms** in order to anonymize the requests sent to the Location Based Server.

- **Mix Zones**
  - Two different areas:
    - **Application zones**
      - Areas in which users can issue requests to LBS.
    - **Mix zones**:
      - Areas in which users cannot issue requests; however, they get a new pseudonym after entering in the area.
Privacy & Location-awareness

- **Pseudonym-based mechanisms**
  - Each user is using *pseudonyms* in order to anonymize the requests sent to the Location Based Server.

- Here device 1 is using the pseudonym X (Application Zone), and it can issues its requests to the LBS.

---

A. R. Beresford and F. Stajano, Location Privacy in Pervasive Computing, Pervasive Computing, 2003
Privacy & Location-awareness

- **Pseudonym-based mechanisms**
  - Each user is using *pseudonyms* in order to anonymize the requests sent to the Location Based Server.

었습니다. Mix Zone: it cannot contact the LBS, however it obtains a new pseudonym (Y) from the anonymity server.
Privacy & Location-awareness

Pseudonym-based mechanisms

- Each user is using **pseudonyms** in order to anonymize the requests sent to the Location Based Server.

- When entering in a new Applicazione Zone (even the previous one), the Device 1 is now using the pseudonym Y.
Privacy & Location-awareness

- Perturbation-based mechanisms

  - The data sent from the mobile nodes (or from the trusted servers) are manipulated by adding some noise.

    CLOAK DROID

    - Simply truncate the GPS coordinates, based on discrete cell grid of size $s_{lat}$, $s_{long}$.

    - $\text{lat}' = s_{lat}\lfloor\text{lat}/s_{lat}\rfloor$  \hspace{1cm} $\text{long}' = s_{long}\lfloor\text{long}/s_{long}\rfloor$

    - Tested across six popular Android locations, and on five different urban scenarios.

<table>
<thead>
<tr>
<th>Name</th>
<th>Objects in List</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gasbuddy</td>
<td>Gas stations</td>
</tr>
<tr>
<td>Restaurant Finder</td>
<td>Restaurants</td>
</tr>
<tr>
<td>Hospitals Near Me</td>
<td>Hospitals</td>
</tr>
<tr>
<td>WebMD</td>
<td>Pharmacies and clinics</td>
</tr>
<tr>
<td>Walmart</td>
<td>Stores</td>
</tr>
<tr>
<td>TD Bank</td>
<td>ATMs and branches</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>City</th>
<th>Population</th>
<th>Radius (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York, NY</td>
<td>8,200,000</td>
<td>30</td>
</tr>
<tr>
<td>Dallas, TX</td>
<td>1,200,000</td>
<td>20</td>
</tr>
<tr>
<td>New Haven, CT</td>
<td>130,000</td>
<td>10</td>
</tr>
<tr>
<td>Baltimore, MD</td>
<td>620,000</td>
<td>12</td>
</tr>
<tr>
<td>Redmond, WA</td>
<td>54,000</td>
<td>4</td>
</tr>
<tr>
<td>Decatur, TX</td>
<td>6,000</td>
<td>4</td>
</tr>
</tbody>
</table>

Privacy & Location-awareness

- Perturbation-based mechanisms
  - The data sent from the mobile nodes (or from the trusted servers) are manipulated by adding some noise.

---

Privacy & Location-awareness

- Definitions
- Privacy Threats
- Privacy Architectures
- Privacy models
- Privacy techniques
- Open issues

Source: https://www.techlicious.com/

Privacy & Location-awareness

- New protection mechanisms
- Quantify location privacy
- Availability of datasets for testing purposes
- Users awareness of privacy risks
- Implementation of privacy mechanisms on real devices.