Subtleties of 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."


When:
- For D-Day attack: troop location privacy, not important 70 years later
- Alert fires to tell your family you stop for pancakes.

How:
- "Michael Michtem Chocolates"
- "Weight Watchers"
- "To what extent: Accuracy high enough to distinguish?"

Computational Location Privacy

- Law – Privacy regulations enforced by government
- Policy – Trust-based, often from institutions
- Encryption – Applies to any type of data.

Computational Location Privacy – Exploits geometric nature of data with algorithms
Why Reveal Your Location?

If you want to know your location, sometimes have to tell someone else.

- Loki Wi-Fi locator – send your Wi-Fi fingerprint and get back (lat,long)
- Quova Reverse IP – send your IP address and get back (lat,long)
- UlbSense – static sensors receive UWB to compute (x,y,z)
- Cricket – MIT
- POLS – Intel Research

Exceptions

Variable Pricing

Congestion Pricing

Pay As You Drive (PAYD) Insurance

Traffic Probes

http://dash.net/
Social Applications

- Dodgeball
- Geotagged Flickr
- Geotagged Twitter
- MotionBased

Location-Based Services

- Local Information
- Tracking
- Navigation
- Games
- Location Alerts

Research

- OpenStreetMap (London)
- MSMS (Seattle)
People Don’t Care about Location Privacy

- 74 U. Cambridge CS students
  - Would accept £10 to reveal 28 days of measured locations (£20 for commercial use)¹

- 226 Microsoft employees
  - 14 days of GPS tracks in return for 1 in 100 chance for £200 MP3 player

- 62 Microsoft employees
  - Only 25% trusted on not sharing GPS data outside

- 12 with location-sensitive message service in Seattle
  - Privacy concerns fairly light²

- 55 Finland interviews on location-aware services
  - “It did not occur to most of the interviewees that they could be located while using the service.”³

Documented Privacy Leaks

- How Cell Phone Helped Cops Nail Key Murder Suspect – Secret “Pings” that Gave Bouncer Away
  - New York, NY, March 15, 2006

- Stalker VicVicto Should Check For GPS
  - Milwaukee, WI, February 6, 2003

- A Face Is Exposed for AOL Searcher No. 4417749
  - New York, NY, August 9, 2006

- Real @me celebrity sightings http://www.gawker.com/stalker/
- A Face Is Exposed for AOL Searcher No. 4417749
  - New York, NY, August 9, 2006

Subtleties of Location Privacy

- Interviews of location based services users
  - Less worry about location privacy in closed campus (1)

- Interviews in 5 EU countries
  - Price for location varied depending on intended use (2)

- Greeks significantly more concerned about location privacy
  - Study two months after wiretapping of Greek politicians (2)
Computational Location Privacy Threats

- Not computational: stalking, spying, peeping
- Not computational: browsing geocoded images
- Not computational: browsing GPS tracks

Significant Locations From GPS Traces

- ContextMent (Marmasse & Schmandt, 2000)
  - Consistent loss of GPS signal → salient location
  - User gives label (e.g., "Grandma’s")
- Project Lachesis (Hariharan & Toyama, 2004)
  - Time/space clustering
  - Hierarchical
- Common aim: find user’s significant locations, e.g. home, work

Context Inference

- Project Lachesis (Kang, Welbourne, Stewart, & Borriello, 2004)
  - Time-based clustering of GPS (lat,long)

- Common aim: find user’s significant locations, e.g. home, work
Context Inference - Wow

Machine learning to infer these properties based only on time-stamped location history.

Location is Quasi-Identifier

Quasi-Identifier – "their values, in combination, can be linked with external information to reidentify the respondents to whom the information refers. A typical example of a single-attribute quasi-identifier is the Social Security Number, since knowing its value and having access to external sources it is possible to identify a specific individual.”

Simulated Location Privacy Attack 1

Experiment
- Attach pseudonym to each person's location history
- Check
  - Where does person spend majority of time?
  - Who spends most time at any given desk?
- Found correct name of all participants
Simulated Location Privacy Attack 2

Experiment:
- GPS histories from 65 drivers
- Cluster points at stops
- Homes are clusters 4 p.m. – midnight
- Found plausible homes of 85%

Simulated Location Privacy Attack 3

MapPoint Web Service reverse geocoding
- Home Location (84 meters)
- Home Address (52%)
- Identity (5%)

Windows Live Search reverse white pages

Simulated Location Privacy Attack 4

From “multi-target tracking” algorithms originally designed for military tracking

- Three GPS traces with no ID or pseudonym
- Successful data association from physical constraints
Simulated Location Privacy Attack 5

- Home with three occupants
- Two-state sensors
- Continuity analysis on thousands of sensor readings
- 85% correct data association

Permissive, 2005

---

Simulated Location Privacy Attack 6

Refinement operators for working around obfuscated location data

Example refinement sources
- Must stay on connected graph of locations
- Movements are goal-directed
- Maximum speed constant

GIScience 2006

---

Where Do You Want to Go Today?

We already know, more or less.

- Past to destination
- Previous destinations
- Ground cover
- Efficient driving
- Trip time

Efficient driving likelihood

Accuracy = 2 km median error at halfway point of trip
How Do You Want to Get There?

Full Route Prediction

Computational Countermeasures
Computational Countermeasures: Pseudonyms

- **Pseudonymity**
  - Replace owner name of each point with untraceable ID
  - One unique ID for each owner

**Example**
- "Larry Page" → "yellow"
- "Bill Gates" → "red"

- Beresford & Stajano (2003) propose frequently changing pseudonym
- Gruteser & Ishi (2005) showed "multi-target tracking" techniques defeat complete anonymity

Computational Countermeasures: k-Anonymity

- k-anonymity introduced for location privacy by Gruteser & Grunwald, 2003
- They note that temporal ambiguity also gives k-anonymity
- Pattern of service requests could break k-anonymity (Belatti, Wang, Jagoda 2005)

Computational Countermeasures: Mix Zones

- New, unused pseudonym given when user is between "application zones"
- "k-anonymous", when you can be confused with k-1 other people
- Anonymity (i.e. k) varies with busyness of mix zone
- Attack by trying to list all pseudonyms given to a person
- Can use probabilistic paths to associate pseudonyms
Computational Countermeasures: False Reports

• Mix true location report with multiple false reports
• Act only on response from true report

Pervasive Services, 2005

• Communication overhead (addressed in paper)
• Attack by finding most sensible sequence of location reports
• Counter by making false sequences sensible (addressed in paper)

Computational Countermeasures: Obfuscation

A Formal Model of Obfuscation and Negotiation for Location Privacy

• Formalize obfuscation techniques
• Client & server can negotiate what needs to be revealed for successful location based service

Pervasive 2005

original low accuracy low precision
(from Krumm 2007)

Computational Countermeasures: Obfuscation

Conclusion: need lots of obfuscation to counter privacy attack
Computational Countermeasures: Obfuscation

Confuse the multi-target tracker by perturbing paths so they cross.

SECURECOMM 2005

Figure 2: Two users move in parallel. The Path Perturbation algorithm perturbs the parallel segment into a crossing segment.

Conclusion

• Why reveal your location?
  – Lots of good reasons
    – Including just to know your own location
  – Not as much as we might expect
• Computational location privacy threats
  – Lots of sophisticated threats
• Location prediction
  – Even possible to infer your future locations
• Computational countermeasures
  – Much work on countermeasures
  – More work necessary as more threats come