A Survey of Computational Location Privacy

John Krumm
Microsoft Research
Redmond, WA USA

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."

Duckham, M. and L. Kulik, *Location privacy and location-aware computing*, in *Dynamic & Mobile GIS: Investigating Change in Space and Time*, J. Drummond, et al., Editors. 2006, CRC Press: Boca Raton, FL USA. p. 34-51.

When: For D-Day attack, troop location privacy not important 60 years later

How: Alert fires to tell your family whenever you stop for pancakes

"Michael Mischers 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

Outline

- Why reveal your location?
- Do people care about location privacy?
- Computational location privacy threats
- Computational countermeasures
- Quantifying location privacy
- Research issues

Why Reveal Your Location?

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

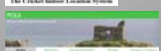
Loki Wi-Fi locator – send your Wi-Fi fingerprint and get back (lat,long)

Quova Reverse IP – send your IP address and get back (lat,long)

UbiSense – static sensors receive UWB to compute (x,y,z)

Exceptions

Cricket – MIT



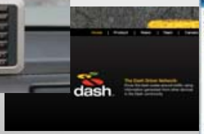
POLS – Intel Research

Variable Pricing

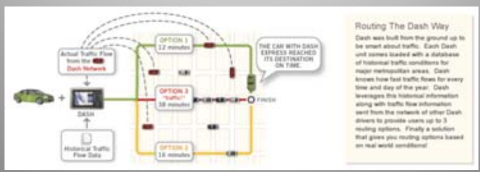
Congestion Pricing

Pay As You Drive (PAYD) Insurance

Traffic Probes

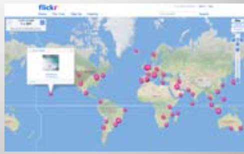



http://dash.net/



Social Applications

Dodgeball



Geotagged Flickr

Geotagged Twitter

MotionBased

Location-Based Services

Navigation

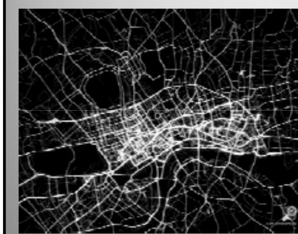
Local Information

Tracking

Games

Location Alerts

Research



OpenStreetMap (London)


MSMLS (Seattle)

Outline

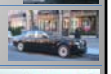
- Why reveal your location?
- Do people care about location privacy?
- Computational location privacy threats
- Computational countermeasures
- Quantifying location privacy
- Research issues

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 21% insisted on not sharing GPS data outside



- 11 with location-sensitive message service in Seattle
- Privacy concerns fairly light (2)

- 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)

(1) Danezis, G., S. Lewis, and R. Anderson. *How Much is Location Privacy Worth? in Fourth Workshop on the Economics of Information Security.* 2005. Harvard University.

(2) Iachello, G., et al. *Control, Deception, and Communication: Evaluating the Deployment of a Location-Enhanced Messaging Service.* in *UbiComp 2005: Ubiquitous Computing.* 2005. Tokyo, Japan.

(3) Kaasinen, E., *User Needs for Location-Aware Mobile Services. Personal and Ubiquitous Computing*, 2003. 7(1): p. 70-79.

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 Victims Should Check For GPS
Milwaukee, WI, February 6, 2003

Real time 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)]

[(1)] Barkhuus, L., *Privacy in Location-Based Services, Concern vs. Coolness*, in *Workshop on Location System Privacy and Control, Mobile HCI 2004*. 2004: Glasgow, UK.

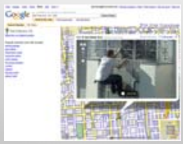
[(2)] Cvrček, D., et al., *A Study on The Value of Location Privacy*, in *Fifth ACM Workshop on Privacy in the Electronic Society*. 2006, ACM: Alexandria, Virginia, USA. p. 109-118.

Outline

- Why reveal your location?
- Do people care about location privacy?
- **Computational location privacy threats**
- Computational countermeasures
- Quantifying location privacy
- Research issues

Computational Location Privacy Threats

Not computational: stalking, spying, peeping



Not computational: browsing geocoded images

Not computational: browsing GPS tracks

Significant Locations From GPS Traces

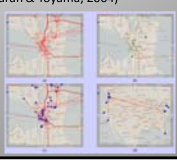
commotion (Marmasse & Schmandt, 2000)
- consistent loss of GPS signal → salient location
- user gives label (e.g. "Grandma's")

Ashbrook & Starner, 2003
- cluster places with lost GPS signal
- user gives label

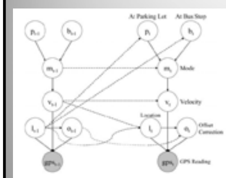
Common aim: find user's significant locations, e.g. home, work

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

Project Lachesis (Hariharan & Toyama, 2004)
- time/space clustering
- hierarchical



Context Inference



Patterson, Liao, Fox & Kautz, 2003
- GPS traces
- Infer mode of transportation (bus, foot, car)
- Route prediction

Location says a lot about you

Krumm, Letchner & Horvitz, 2006
- Noisy GPS matched to road driven
- Constraints from speed & road connectivity

Predestination (Krumm & Horvitz, 2006)
- Predict destination
- Extends privacy attack into future

Context Inference - Wow


Inferring Long-term User Properties based on Users' Location History

IJCAI 2007

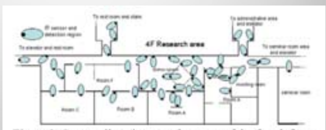


Figure 3: Sensor allocation map for a part of the fourth floor.

Indoor location sensors

Table 1: User properties.

user property	range
AGE	under24, 24-29, 30-34, 35-39, over40
POSITION	ac*, full-time researcher, part-time researcher, technical-staff, temporary-staff
TEAM	research-group-A, -B, -C, -D, secretaries, administrators
WORK-FREQUENCY**	high, middle, low
COFFEE***	high, middle, low
SMOKING	yes, no
ROOM*	A, B, C, D, E, F
COMMUTING**	station-A, station-B

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

Good: TEAM, ROOM
OK: AGE, COFFEE, SMOKING
Bad: POSITION, WORK FREQUENCY

Location is Quasi-Identifier


Protecting Privacy Against Location-based Personal Identification

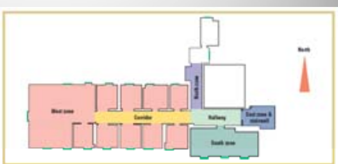
Quasi-Identifier – "their values, in combination, can be linked with external information to reidentify the respondents to whom the information refers. A typical example of a single-attribute quasi-identifier is the Social Security Number, since knowing its value and having access to external sources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

Secure Data Management, VLDB workshop, 2005

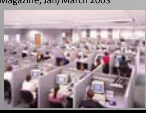
Simulated Location Privacy Attack 1


Location Privacy in Pervasive Computing

IEEE Pervasive Computing Magazine, Jan/March 2003



Active BAT indoor location system



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


Enhancing Security and Privacy in Traffic-Monitoring Systems

IEEE Pervasive Computing Magazine, Oct/Dec 2006

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


Fifth International Conference on Pervasive Computing (Pervasive 2007), May 13-16, Toronto, Ontario, Canada

Inference Attacks on Location Tracks

John Krumm
Microsoft Research

Pervasive 2007

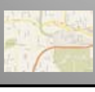
GPS Tracks (172 people)

Home Location (61 meters)

Home Address (12%)

Identity (5%)

MapPoint Web Service reverse geocoding



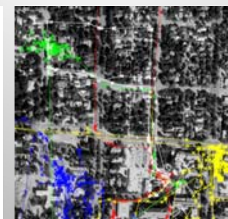
Windows Live Search reverse white pages

Simulated Location Privacy Attack 4


On the Anonymity of Periodic Location Samples

Marco Gruteser and Baik Hoh

Security in Pervasive Computing, 2005

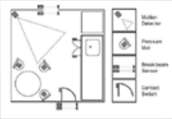
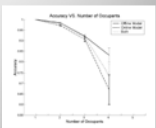


- Three GPS traces with no ID or pseudonym
- Successful data association from physical constraints

From "multi-target tracking" algorithms originally designed for military tracking

Simulated Location Privacy Attack 5


Simultaneous Tracking & Activity Recognition (STAR) Using Many Anonymous, Binary Sensors

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

Pervasive, 2005

Simulated Location Privacy Attack 6

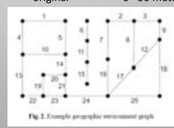

A spatiotemporal model of strategies and counter strategies for location privacy protection

GIScience 2006

Refinement operators for working around obfuscated location data


original


σ= 50 meters noise added

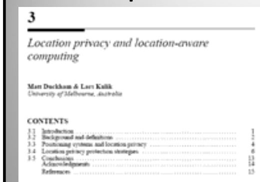


Example refinement sources
- Must stay on connected graph of locations
- Movements are goal-directed
- Maximum speed constraint

Outline

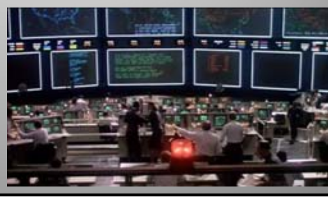
- Why reveal your location?
- Do people care about location privacy?
- Computational location privacy threats
- Computational countermeasures
- Quantifying location privacy
- Research issues

Computational Countermeasures


3
Location privacy and location-aware computing
Matt Duckham & Lars Kulik

Dynamic & Mobile GIS: Investigating Change in Space and Time, CRC Press, 2006

Four ways to enhance location privacy
1. Regulations – govt. enforced
2. Policies – trust-based agreements
3. Anonymity – pseudonyms and/or ambiguity
4. Obfuscation – reduce quality of data



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 & Hoh (2005) showed "multi-target tracking" techniques defeat complete anonymity

Computational Countermeasures: k-Anonymity

I'm chicken # 341, and I'm in this building (along with k-1 other chickens).

I'm chicken # 341, and I visited this place in the past 21 minutes (along with k-1 other chickens).

- 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 (Bettini, Wang, Jajodia 2005)

Computational Countermeasures: Mix Zones

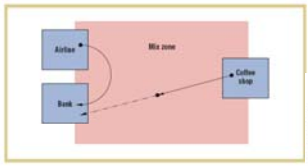
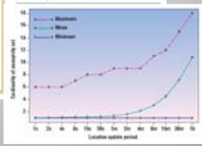


Figure 1. A sample mix zone arrangement with three application zones. The airline agency (A) is much closer to the bank (B) than the coffee shop (C). Users leaving A and C at the same time might be distinguishable on arrival at B.



Beresford & Stajano, 2003

- 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

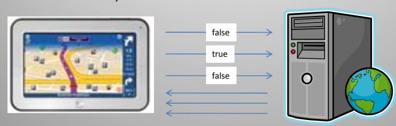
Computational Countermeasures: False Reports


An Anonymous Communication Technique using Dummies for Location-based Services

Hidetoshi Kido† Yutaka Yanagisawa†† Tetsuji Satoh†,††
†Graduate School of Information Science and Technology, Osaka University
††NTT Communication Science Laboratories, NTT Corporation
h-kido@ist.osaka-u.ac.jp yutaka@cslab.kecl.ntt.co.jp satoh.tetsuji@lab.ntt.co.jp

- 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) (fun research project)

Computational Countermeasures: Obfuscation


A Formal Model of Obfuscation and Negotiation for Location Privacy

Matt Duckham[1] and Lars Kulik[2]

[1] Department of Geomatics, University of Melbourne, Victoria, 3010, Australia
[2] Department of Computer Science and Software Engineering, University of Melbourne, Victoria, 3010, Australia

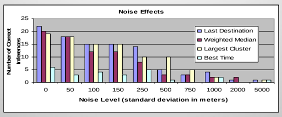
Pervasive 2005

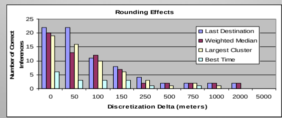
- Formalizes obfuscation techniques
- Client & server can negotiate what needs to be revealed for successful location based service

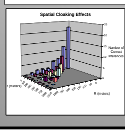
original low accuracy low precision

(from Krumm 2007)

Computational Countermeasures: Obfuscation







Conclusion: need lots of obfuscation to counter privacy attack


Inference Attacks on Location Tracks

Computational Countermeasures: Obfuscation


Protecting Location Privacy Through Path Confusion

Baik Hoh, WINLAB, ECE Department, Rutgers, The State University of New Jersey
Marco Gruteser, WINLAB, ECE Department, Rutgers, The State University of New Jersey

SECURECOMM 2005

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

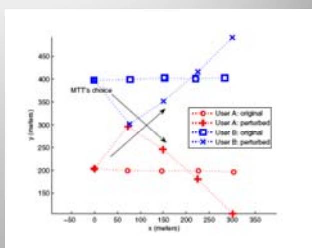
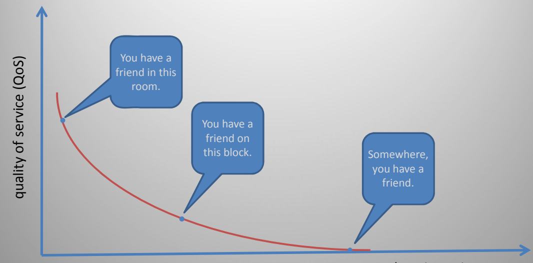


Figure 2. Two users move in parallel. The Path Perturbation algorithm perturbs the parallel segment into a crossing segment.

Outline

- Why reveal your location?
- Do people care about location privacy?
- Computational location privacy threats
- Computational countermeasures
- **Quantifying location privacy**
- Research issues

Quantifying Location Privacy

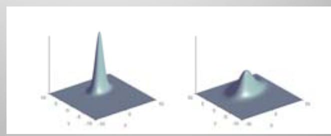


How do you measure the x axis?

Quantifying Location Privacy


Protecting Location Privacy Through Path Confusion

SECURECOMM 2005

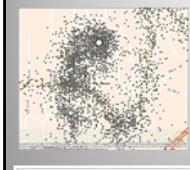


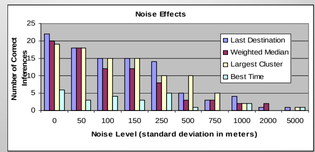
Location privacy = expected error in attacker's estimate of location

- Simple to understand, easy to compute
- Ignores semantics of (lat,long), *e.g.*

 vs. 

Quantifying Location Privacy



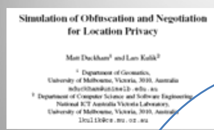



Inference Attacks on Location Tracks

location privacy = accuracy of location data

Same semantic problems as previous

Quantifying Location Privacy


Simulation of Obfuscation and Negotiation for Location Privacy

COSIT 2005

- **You**: I am somewhere in this circle
- **Server**: Given this, I can narrow down the nearest gas station to three possibilities

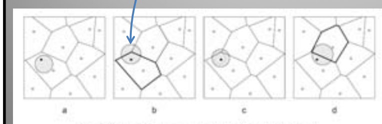
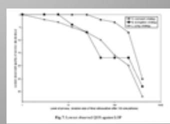


Fig. 1. Negotiation alternatives, illustrated by a Voronoi diagram



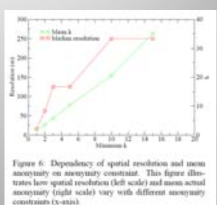
location privacy ~= size of circle

Quantifying Location Privacy


Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking

Mobisys 2003

k-anonymity

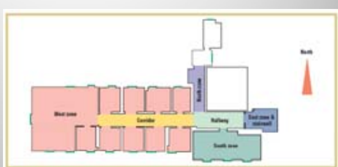


location privacy = k

Quantifying Location Privacy


Location Privacy in Pervasive Computing

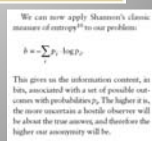
Beresford & Stanjano, IEEE Pervasive Computing Magazine, 2003



location privacy = entropy in location estimate

p(cathedral)	p(Hooters)	entropy
0.5	0.5	1
0.15	0.85	0.60984
0.001	0.999	0.011408



Outline

- Why reveal your location?
- Do people care about location privacy?
- Computational location privacy threats
- Computational countermeasures
- Quantifying location privacy
- Research issues

Research Opportunities

- **Privacy Attitudes** – In the abstract, people don't care. But attitudes depend on many things. What are the dependencies and how strong?
- **Privacy Attacks** – Do one to raise consciousness of problem
- **Inference Attacks** – Find weaknesses in proposed algorithms
- **Hacker Challenge** – Challenge people to break your scheme
- **False Location Reports** – simulate actual motion to make it plausible. Arms race between white hats and black hats.
- **Location Privacy vs. QoS** – Tradeoff location privacy for quality of service
- **Quantify Location Privacy** – find a way that matches perceptions

16th USENIX Security Symposium, 2007

End