Privacy in the Digital Age - CS 5436 / INFO 5303 - Cornell NYC Tech

Location Privacy

Reza Shokri

Location is Identity

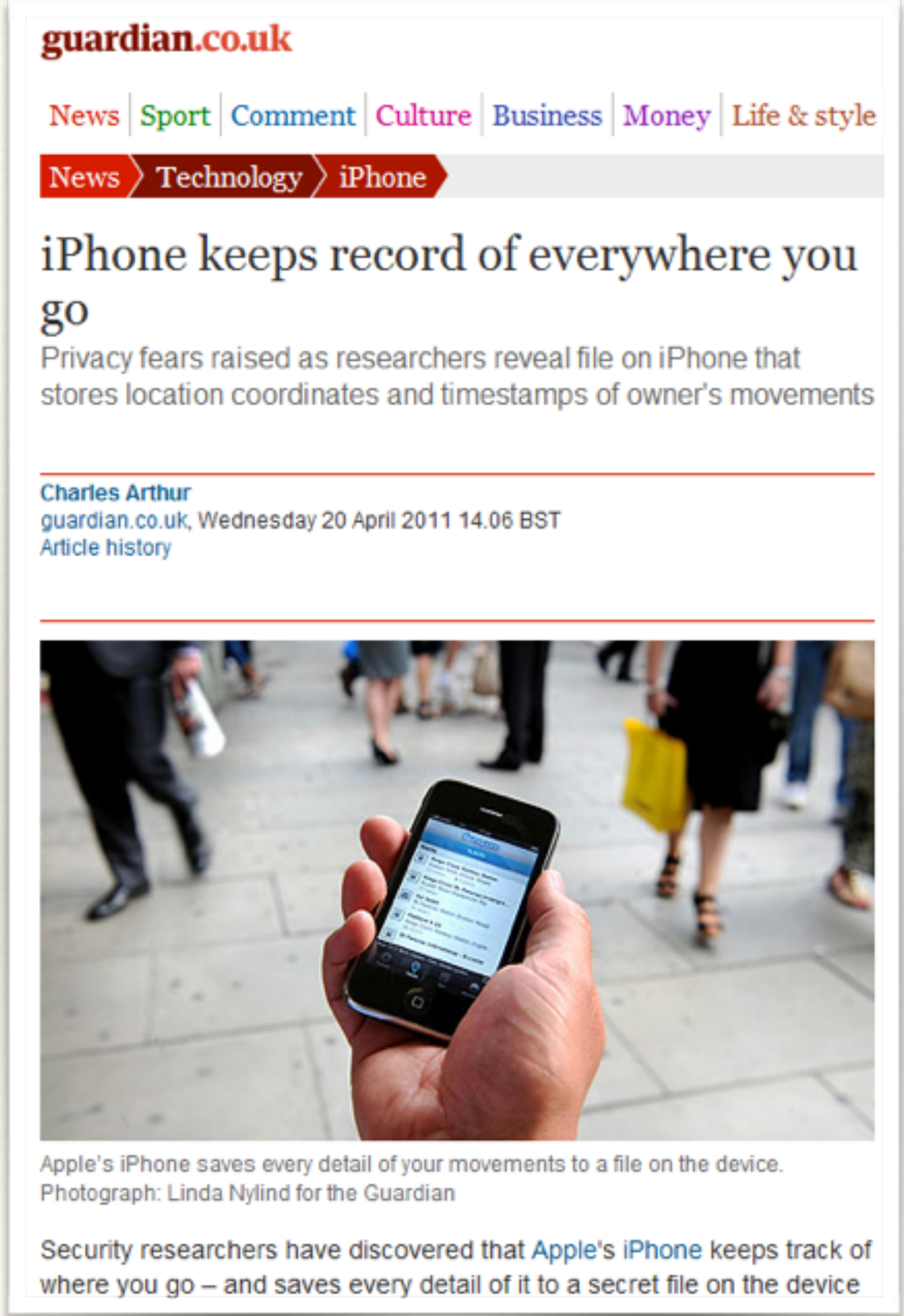
you are where you are

a location trace is not only a set of positions on the map. The contextual information attached to a trace tells much about our habits, interests, activities, and relationships

I Beacon Therefore I Am

- Cellular Networks
- Location-based Services
- Social Networks
- Internet Service Providers
- Wireless Signals
- Car GPS
- E-Pass Cards
- Credit Cards

Presence Disclosure



guardian.co.uk
News | Sport | Comment | Culture | Business | Money | Life & style
News › Technology › iPhone

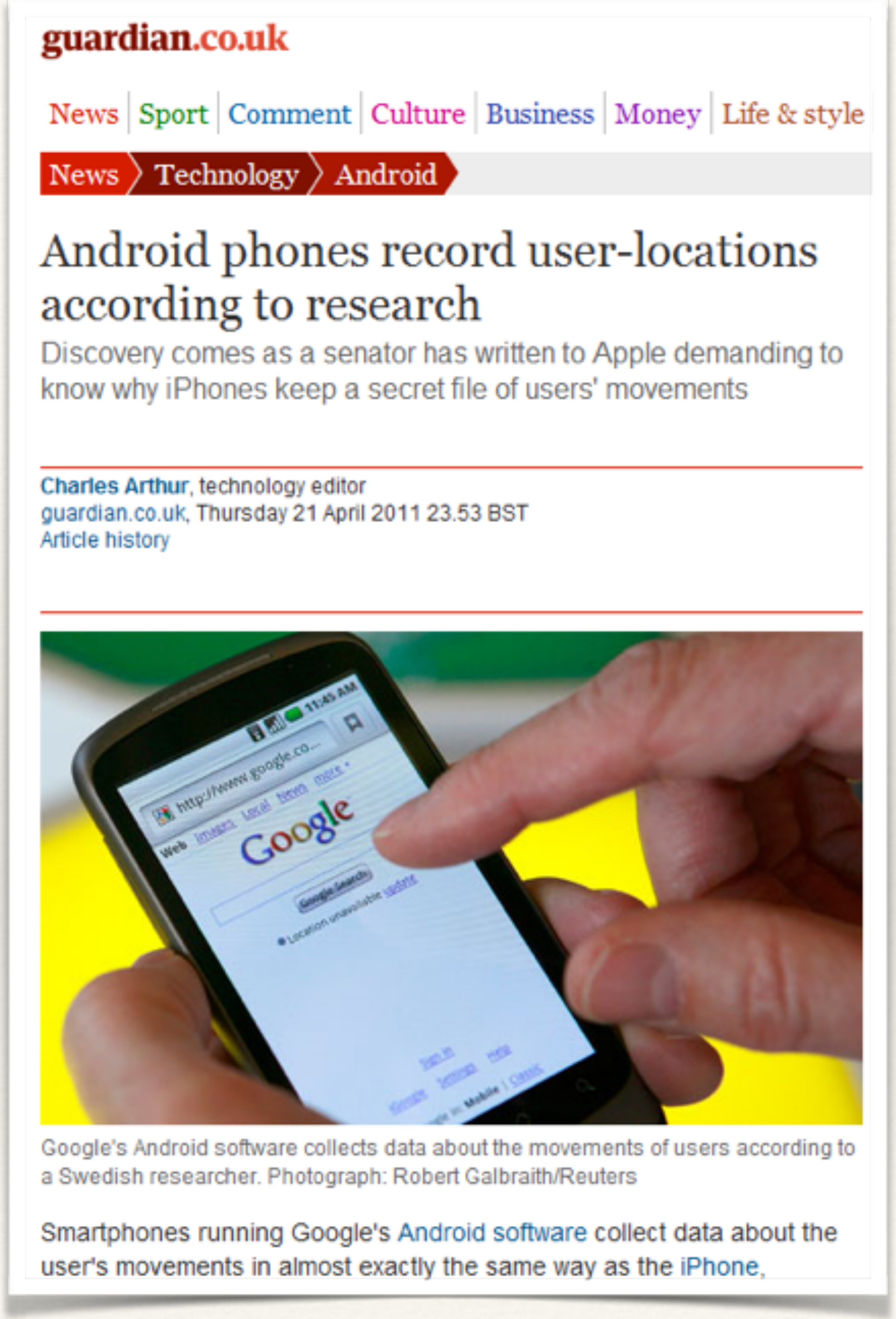
iPhone keeps record of everywhere you go

Privacy fears raised as researchers reveal file on iPhone that stores location coordinates and timestamps of owner's movements

Charles Arthur
guardian.co.uk, Wednesday 20 April 2011 14.06 BST
Article history

Apple's iPhone saves every detail of your movements to a file on the device. Photograph: Linda Nylind for the Guardian

Security researchers have discovered that Apple's iPhone keeps track of where you go – and saves every detail of it to a secret file on the device



guardian.co.uk
News | Sport | Comment | Culture | Business | Money | Life & style
News › Technology › Android

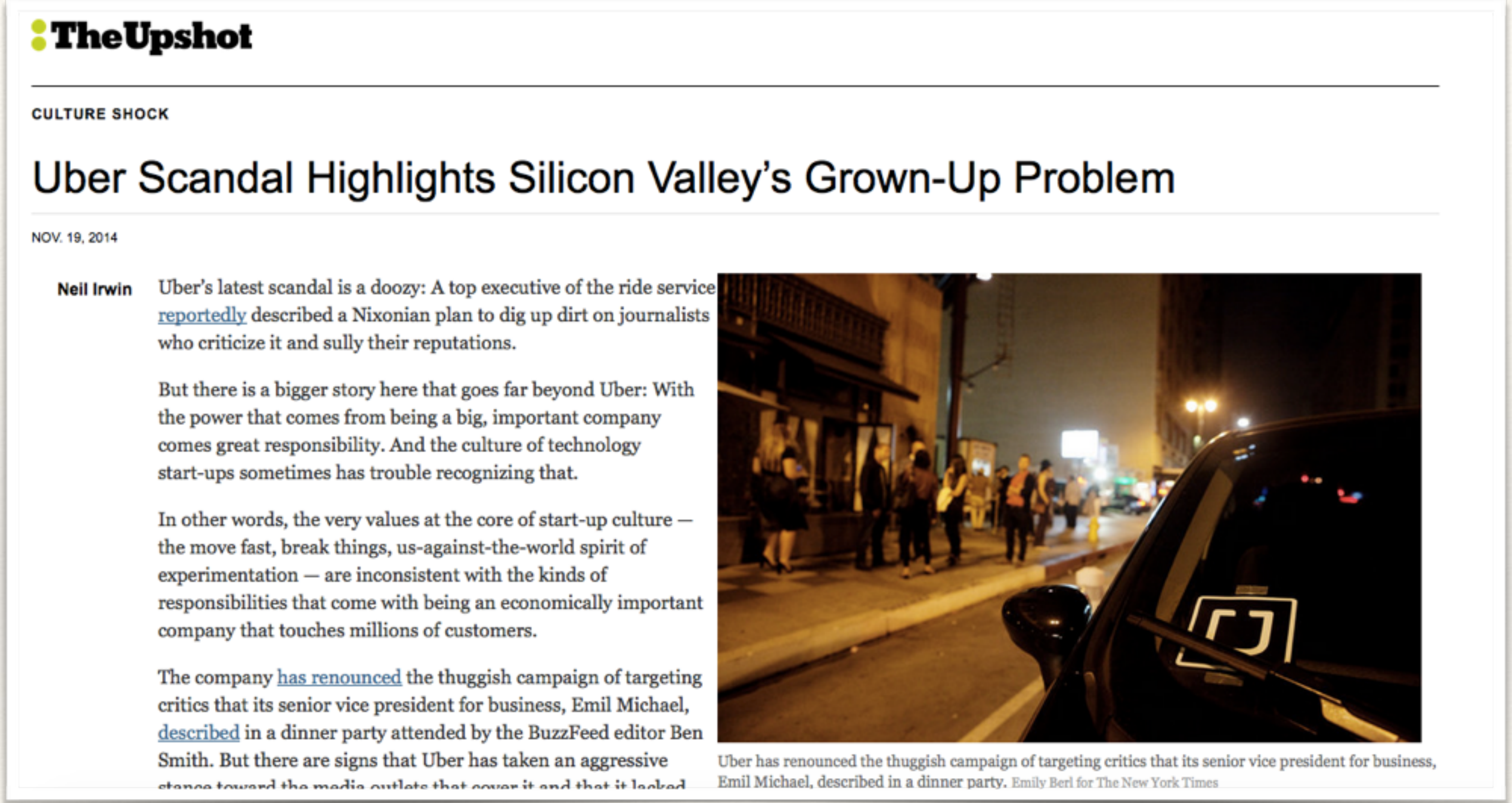
Android phones record user-locations according to research

Discovery comes as a senator has written to Apple demanding to know why iPhones keep a secret file of users' movements

Charles Arthur, technology editor
guardian.co.uk, Thursday 21 April 2011 23.53 BST
Article history

Google's Android software collects data about the movements of users according to a Swedish researcher. Photograph: Robert Galbraith/Reuters

Smartphones running Google's Android software collect data about the user's movements in almost exactly the same way as the iPhone,



TheUpshot

CULTURE SHOCK

Uber Scandal Highlights Silicon Valley’s Grown-Up Problem

NOV. 19, 2014

Neil Irwin

Uber’s latest scandal is a doozy: A top executive of the ride service reportedly described a Nixonian plan to dig up dirt on journalists who criticize it and sully their reputations.

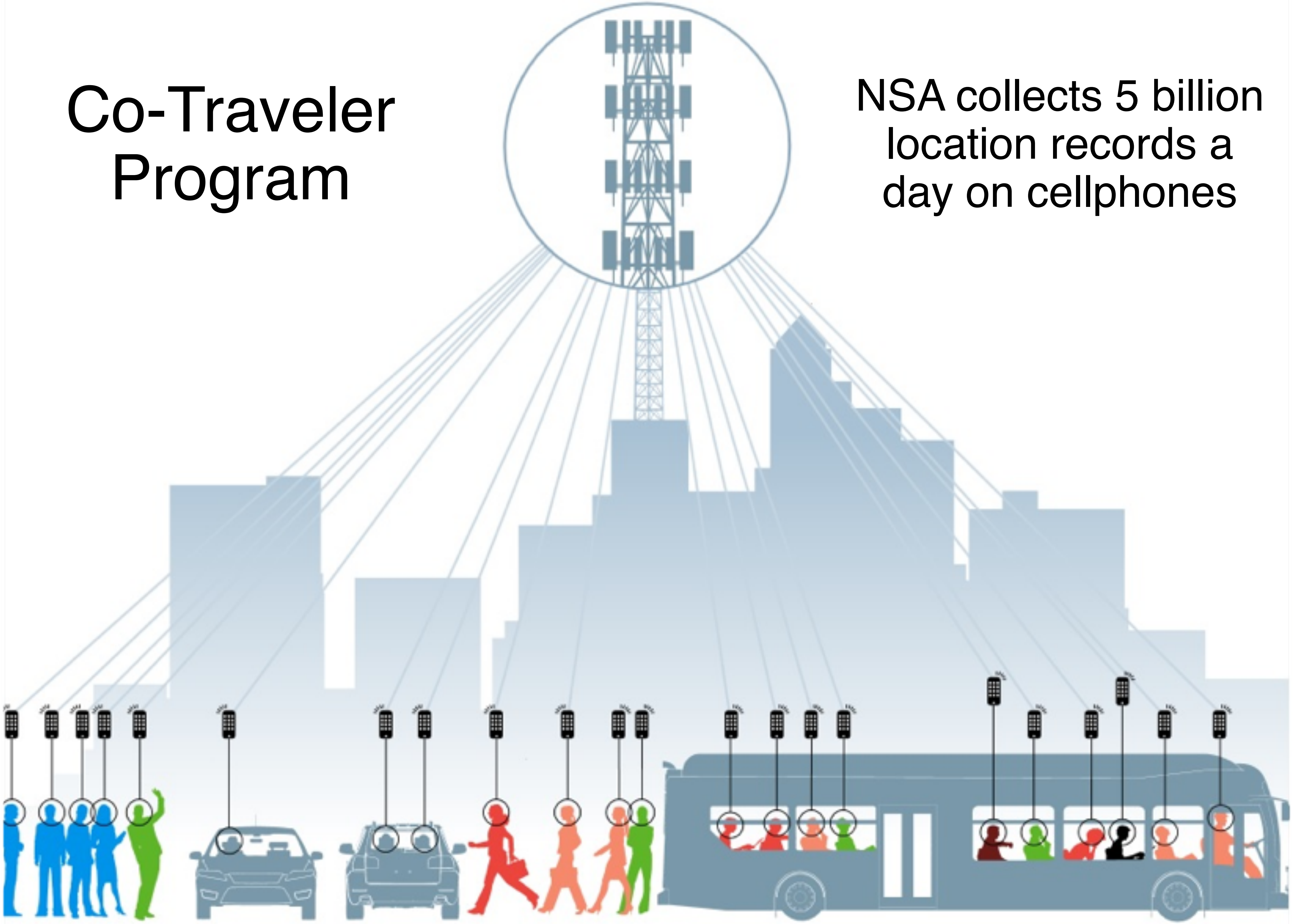
But there is a bigger story here that goes far beyond Uber: With the power that comes from being a big, important company comes great responsibility. And the culture of technology start-ups sometimes has trouble recognizing that.

In other words, the very values at the core of start-up culture — the move fast, break things, us-against-the-world spirit of experimentation — are inconsistent with the kinds of responsibilities that come with being an economically important company that touches millions of customers.

The company has renounced the thuggish campaign of targeting critics that its senior vice president for business, Emil Michael, described in a dinner party attended by the BuzzFeed editor Ben Smith. But there are signs that Uber has taken an aggressive stance toward the media outlets that cover it and that it lacked

Uber has renounced the thuggish campaign of targeting critics that its senior vice president for business, Emil Michael, described in a dinner party. Emily Berl for The New York Times

Co-Traveler Program



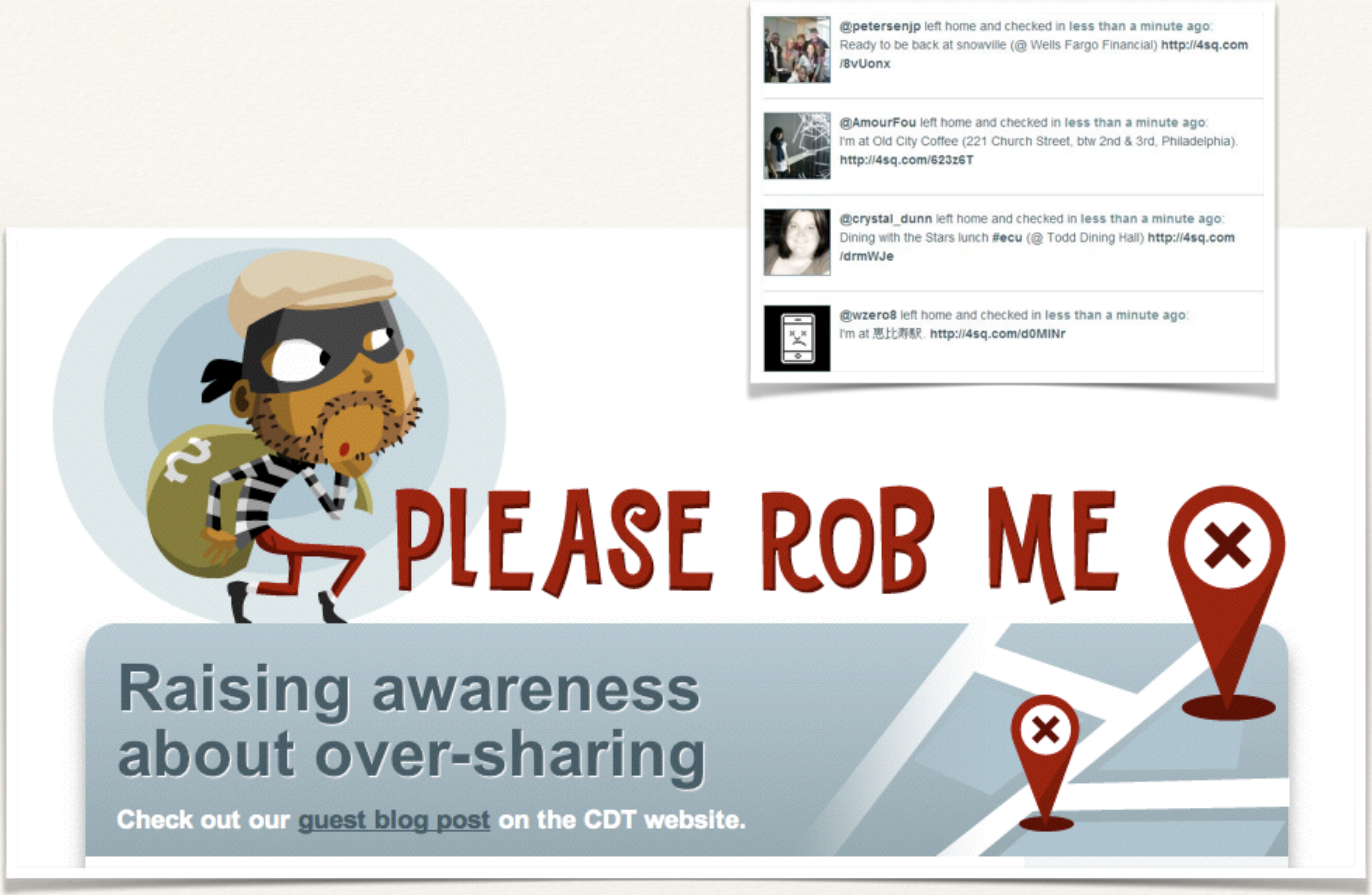
NSA collects 5 billion location records a day on cellphones

washingtonpost.com

Disclosed by Others

- Appear in others' photos
- Checked-in (to a location) by friends

Absence Disclosure



@petersenjp left home and checked in **less than a minute ago**:
Ready to be back at snowville (@ Wells Fargo Financial) **http://4sq.com/8vUonx**

@AmourFou left home and checked in **less than a minute ago**:
I'm at Old City Coffee (221 Church Street, btw 2nd & 3rd, Philadelphia). **http://4sq.com/623z6T**

@crystal_dunn left home and checked in **less than a minute ago**:
Dining with the Stars lunch **#ecu** (@ Todd Dining Hall) **http://4sq.com/drmWJe**

@wzero8 left home and checked in **less than a minute ago**:
I'm at 恵比寿駅. **http://4sq.com/d0MINr**

PLEASE ROB ME

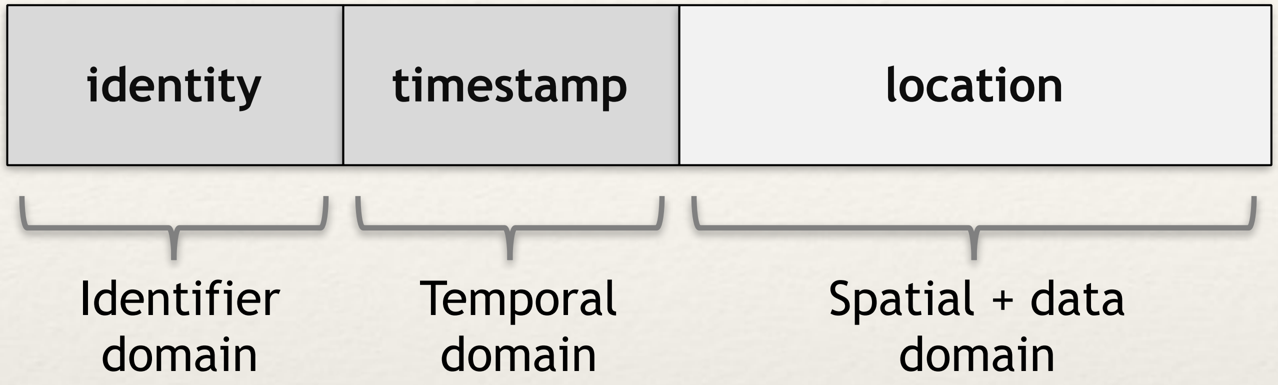
Raising awareness about over-sharing

Check out our guest blog post on the CDT website.

http://pleaserobme.com/

Threats and Countermeasures

Shared Information



identity | timestamp | location

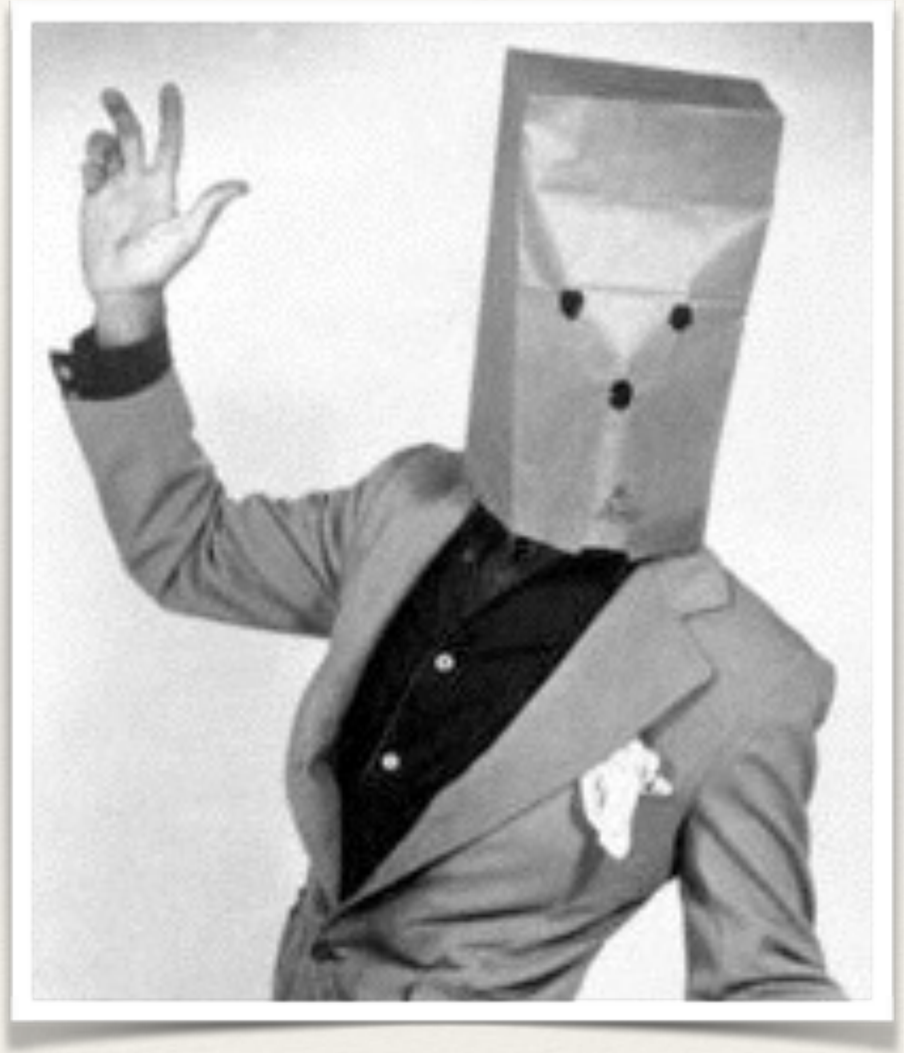
Identifier domain

Temporal domain

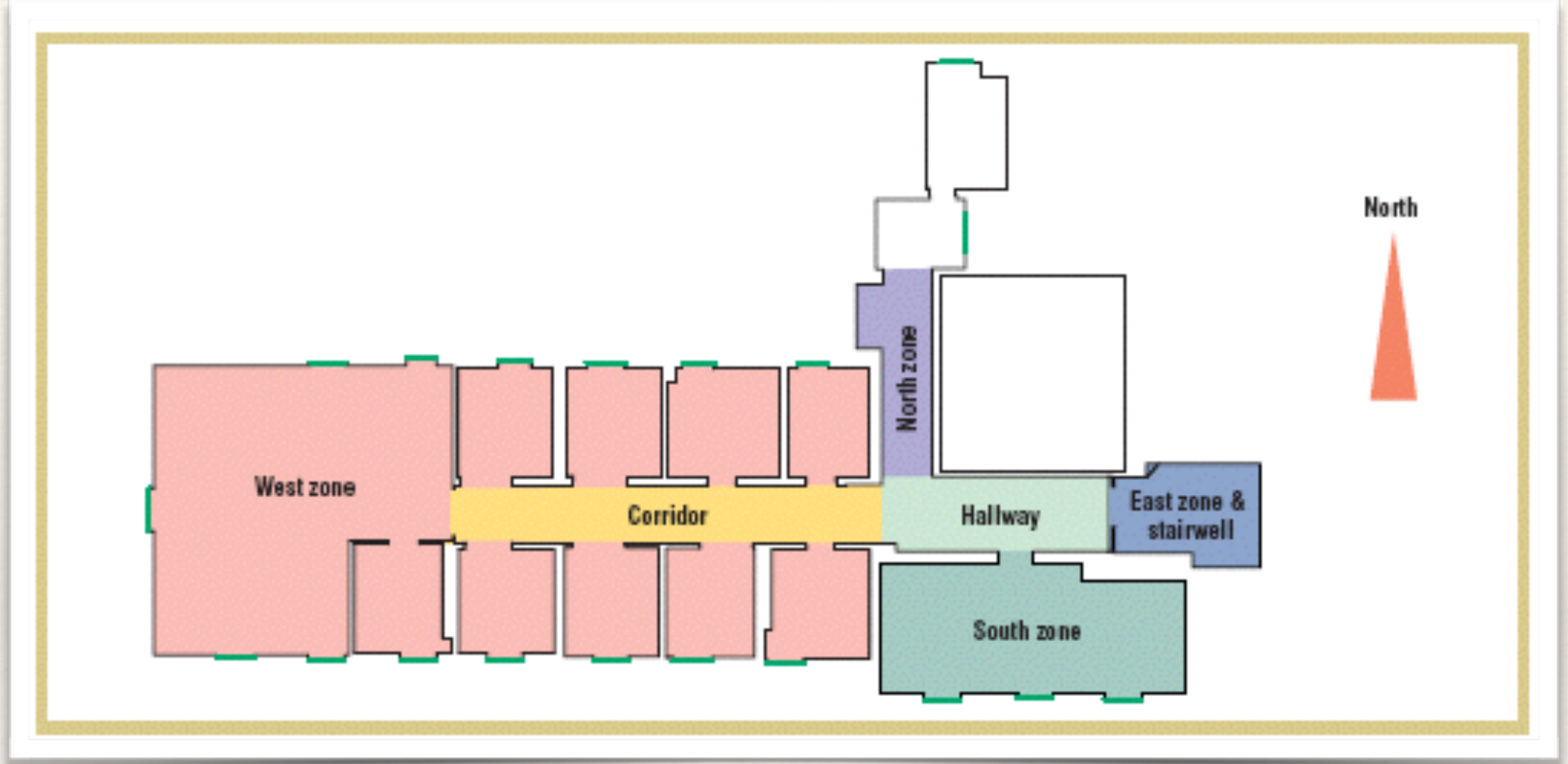
Spatial + data domain

Defense: Use Pseudonym

- To protect your privacy, replace your identity with a fake identity
- Cost?
- Limitations?
- Attack Resilience?



Re-Identification



- Observe locations of anonymized employees in an office environment, and identify people based on their most visited location — All employees identified!

AR. Beresford, F. Stajano, “Location privacy in pervasive computing”, In IEEE Pervasive Computing, 2003

Uniqueness of Significant Locations

- Home and Work locations are pretty unique even at a low granularity location scale

P. Golle and K. Partridge. "On the anonymity of home/work location pairs". In Pervasive, 2009

Uniqueness of Significant Locations

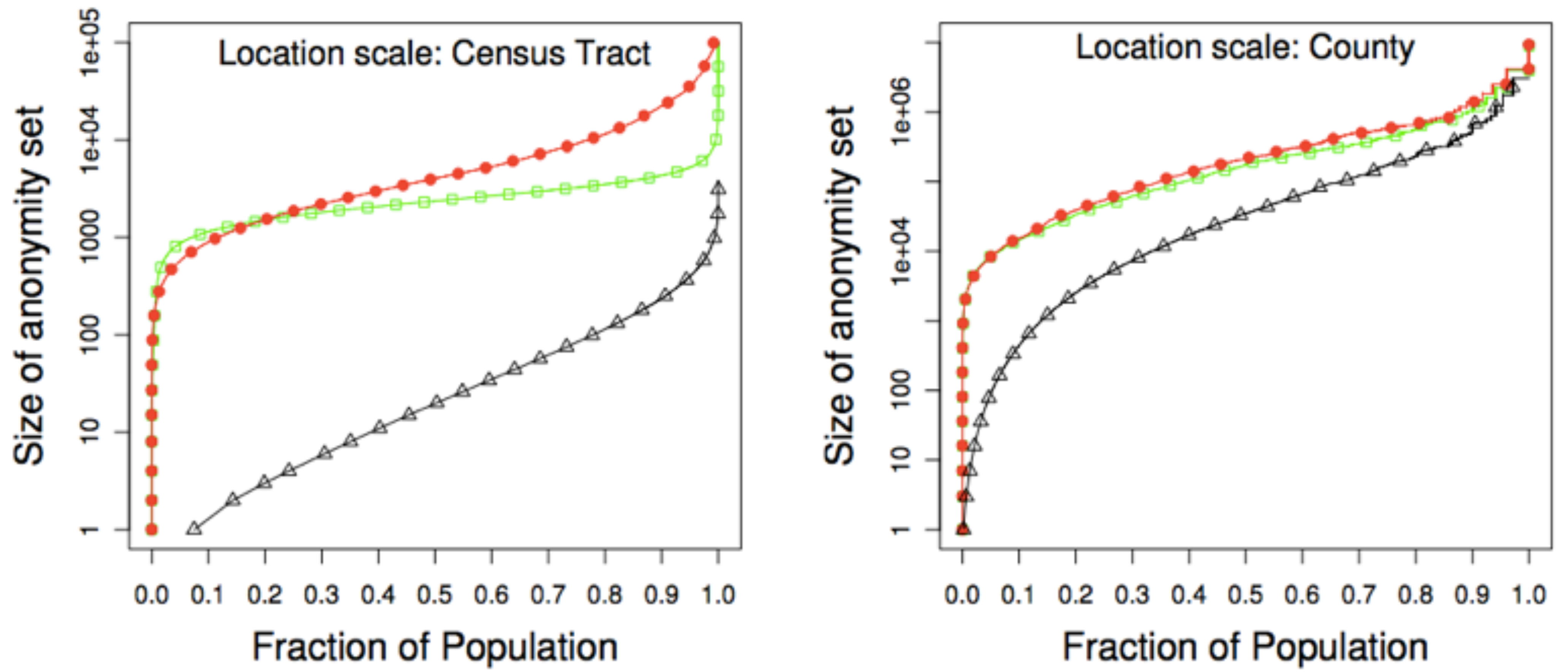


Fig. 1. Size of anonymity set under disclosure of work location (red circles), home location (green squares) or both (black triangles). Location granularity is either census tract (left graph) or county (right graph). Note the different scales on the Y-axes.

- Anonymity Set: Set of indistinguishable individuals

P. Golle and K. Partridge. "On the anonymity of home/work location pairs". In Pervasive, 2009

Predictability of Human Mobility

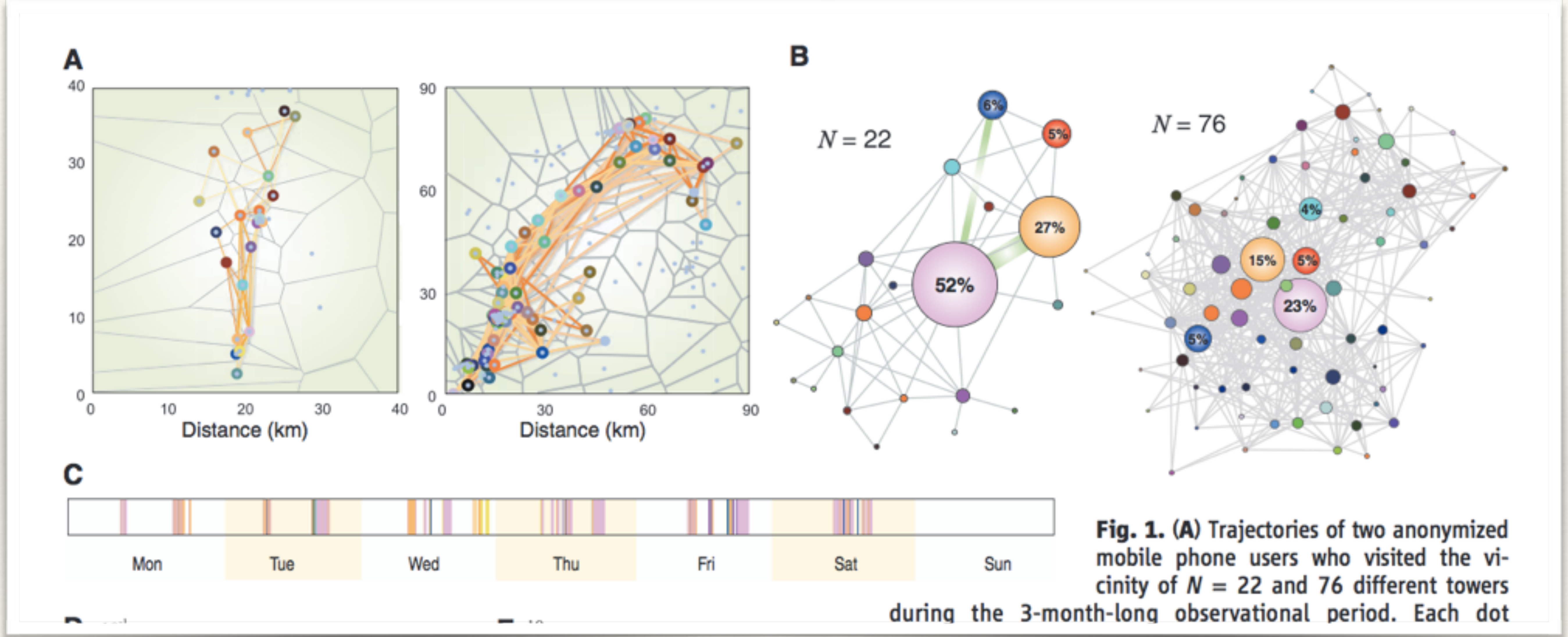
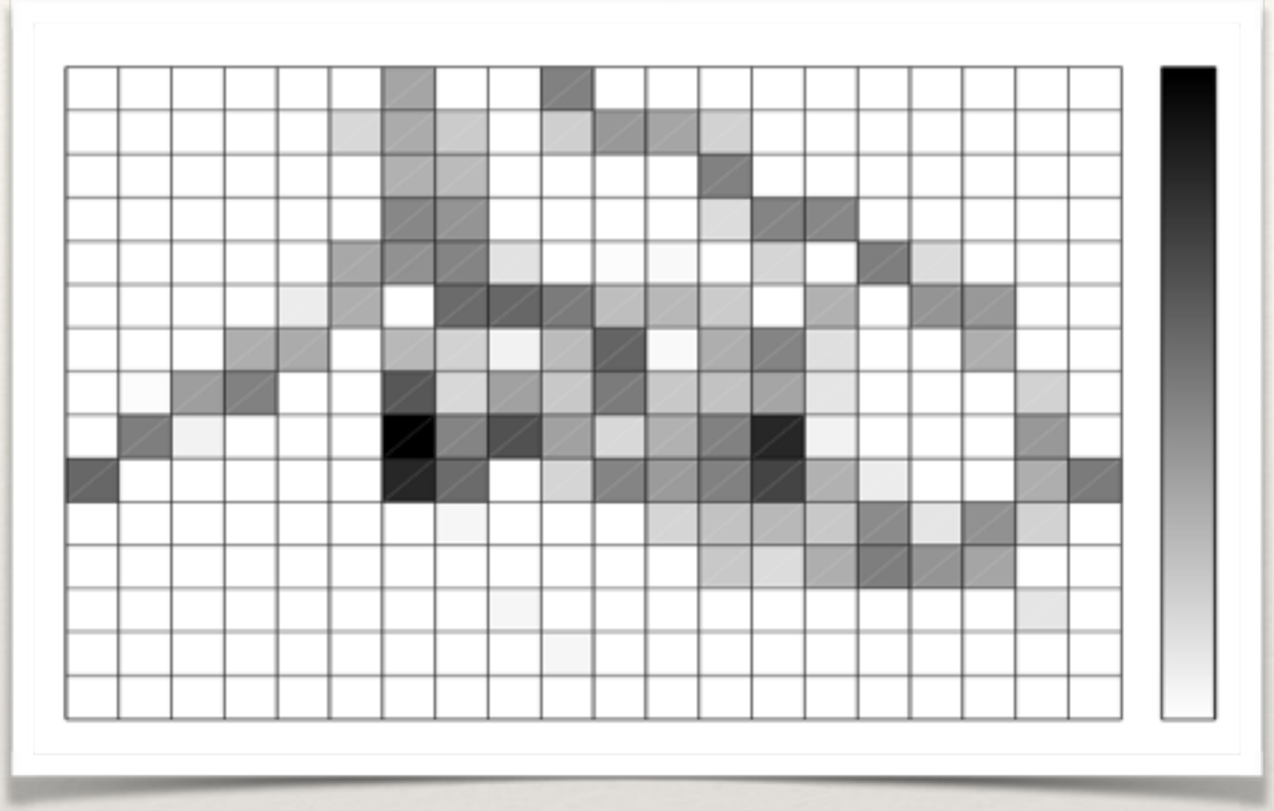


Fig. 1. **(A)** Trajectories of two anonymized mobile phone users who visited the vicinity of *N* = 22 and 76 different towers during the 3-month-long observational period. Each dot

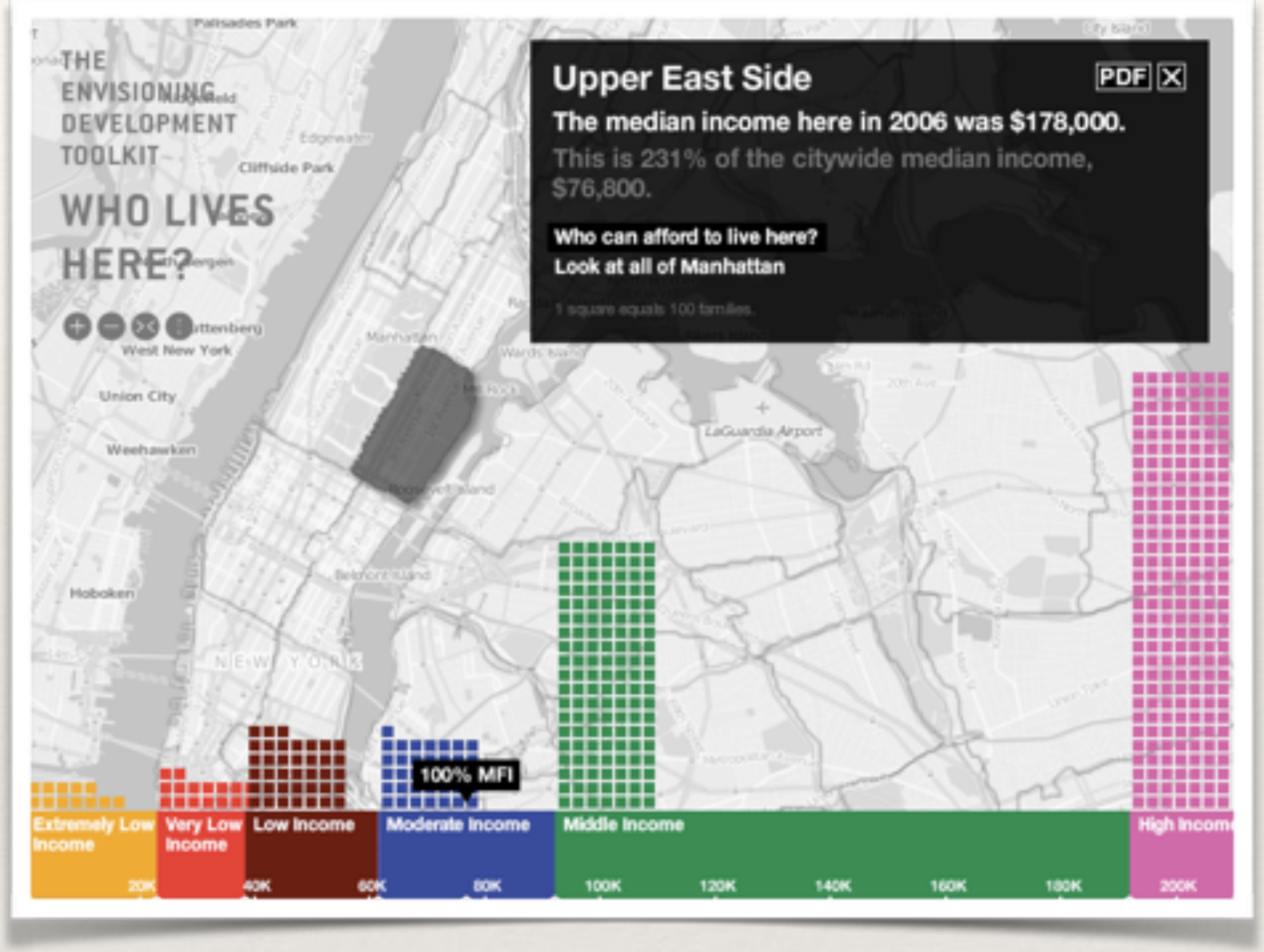
- Humans follow simple predictable location patterns
- Predictability is invariant to the traveled distance

C. Song, Z. Qu, N. Blumm, A. Barabási, "Limits of Predictability in Human Mobility", In Science 2010

Profiling



- Given the significant predictability of human mobility, an adversary can construct a mobility profile of the target that helps him to re-identify or track the target in the future



Upper East Side

The median income here in 2006 was $178,000.

This is 231% of the citywide median income, $76,800.

Who can afford to live here?

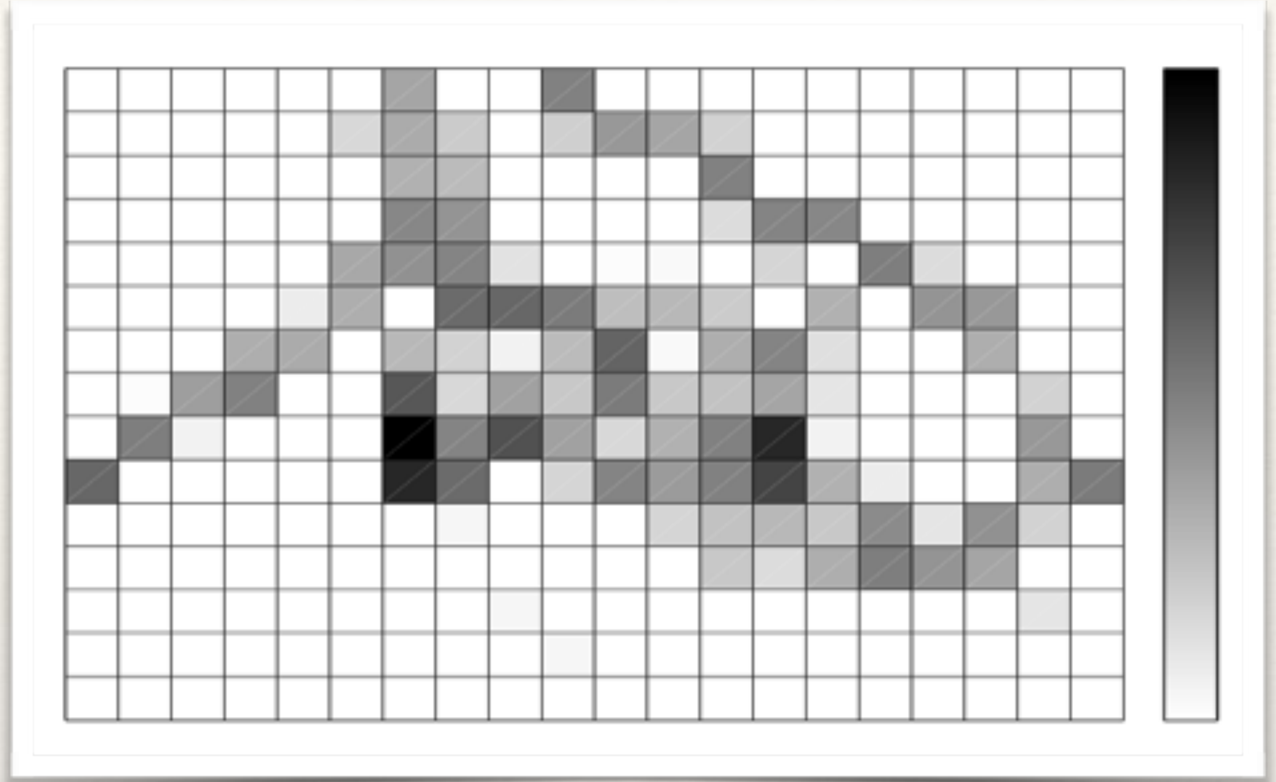
Look at all of Manhattan

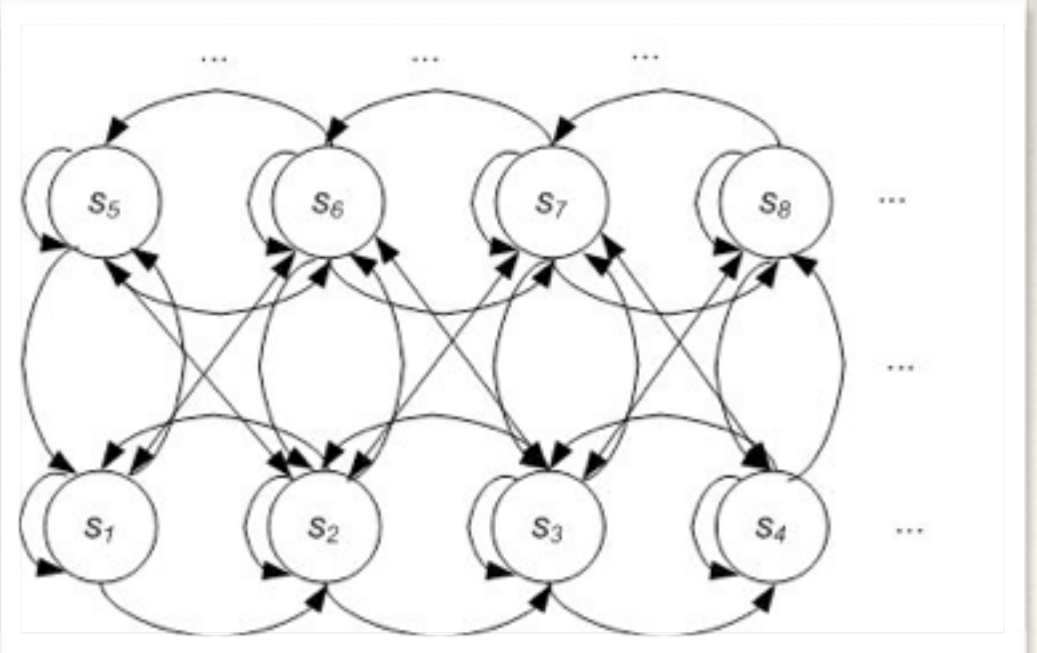
1 square equals 100 families.

- Location profiles reveal information about income, ethnicity, …

A Probabilistic Mobility Profile

- Use Markov Chains to model transition of an individual between different locations
- Each transition is associated with a probability
- Given a location trace, we can learn the probabilities by e.g., normalizing the observed transition counts between locations (maximum likelihood estimation)
- What the adversary knows about the target and uses for constructing target's profile before any attack is referred to as adversary's background knowledge



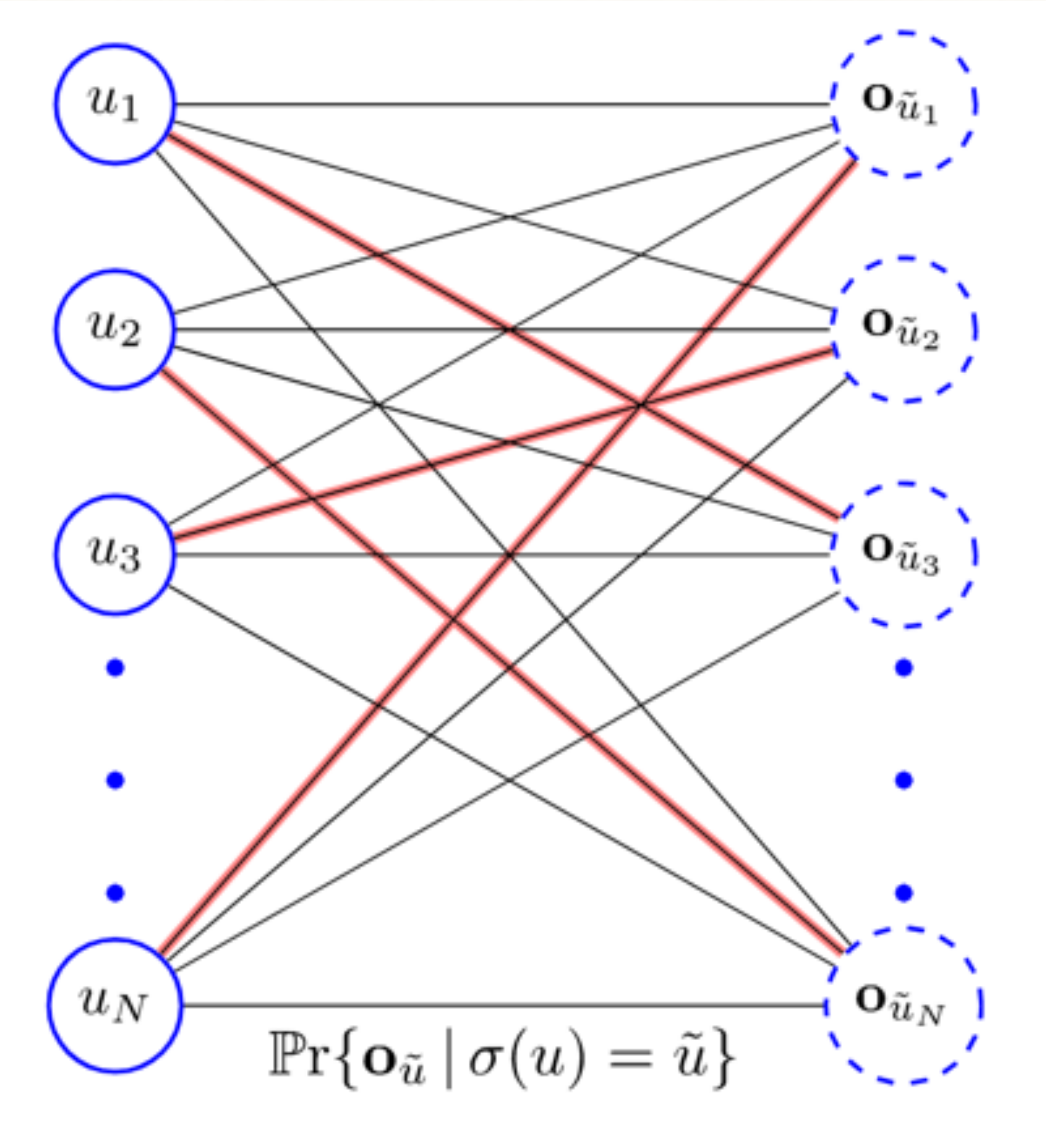


R. Shokri, G. Theodorakopoulos, JY. Le Boudec, JP. Hubaux. "Quantifying Location Privacy", In IEEE Symposium on Security and Privacy, 2011.

Attack: De-Anonymization

- anonymize location traces by removing users’ identities

Users’ Mobility Profiles

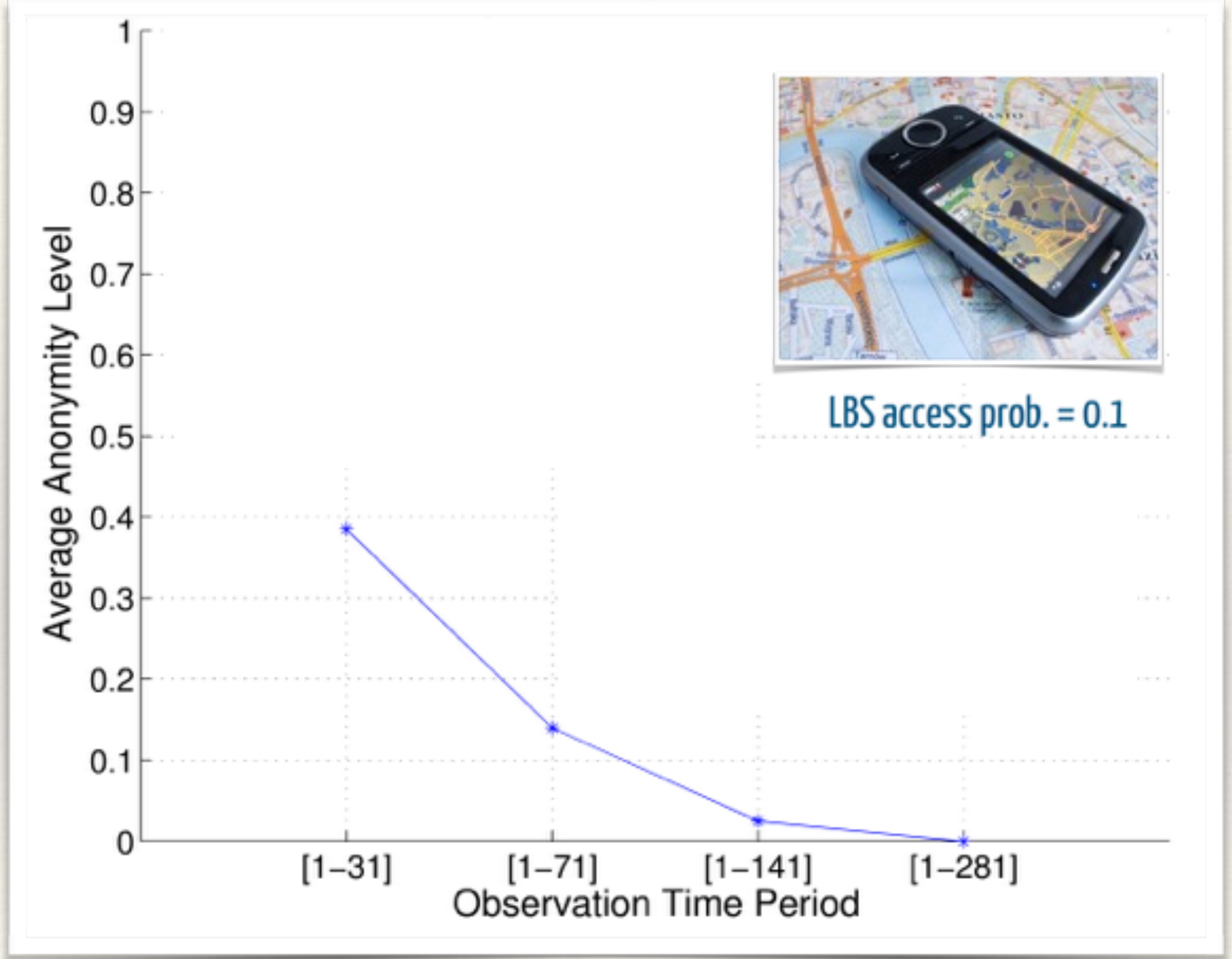

u_1
u_2
u_3
u_N
$\mathbf{o}_{\tilde{u}_1}$
$\mathbf{o}_{\tilde{u}_2}$
$\mathbf{o}_{\tilde{u}_3}$
$\mathbf{o}_{\tilde{u}_N}$
$\mathbb{P}\mathrm{r}\{\mathbf{o}_{\tilde{u}} \,|\, \sigma(u) = \tilde{u}\}$

Anonymous Observed Traces

- compute the probability of each observed trace being generated from each user

- giving this full bipartite graph, compute the most likely assignment of users to traces

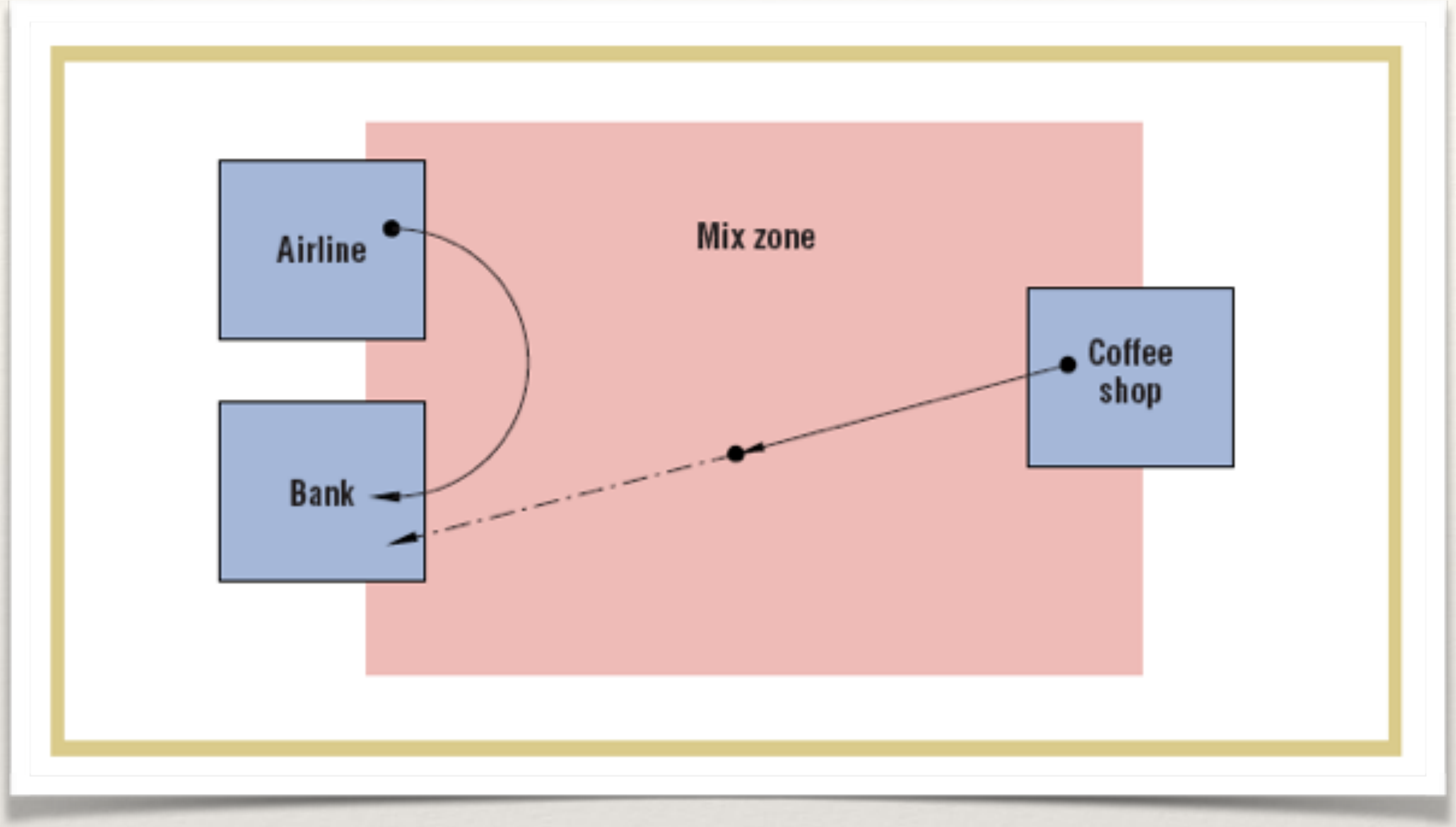
Anonymity



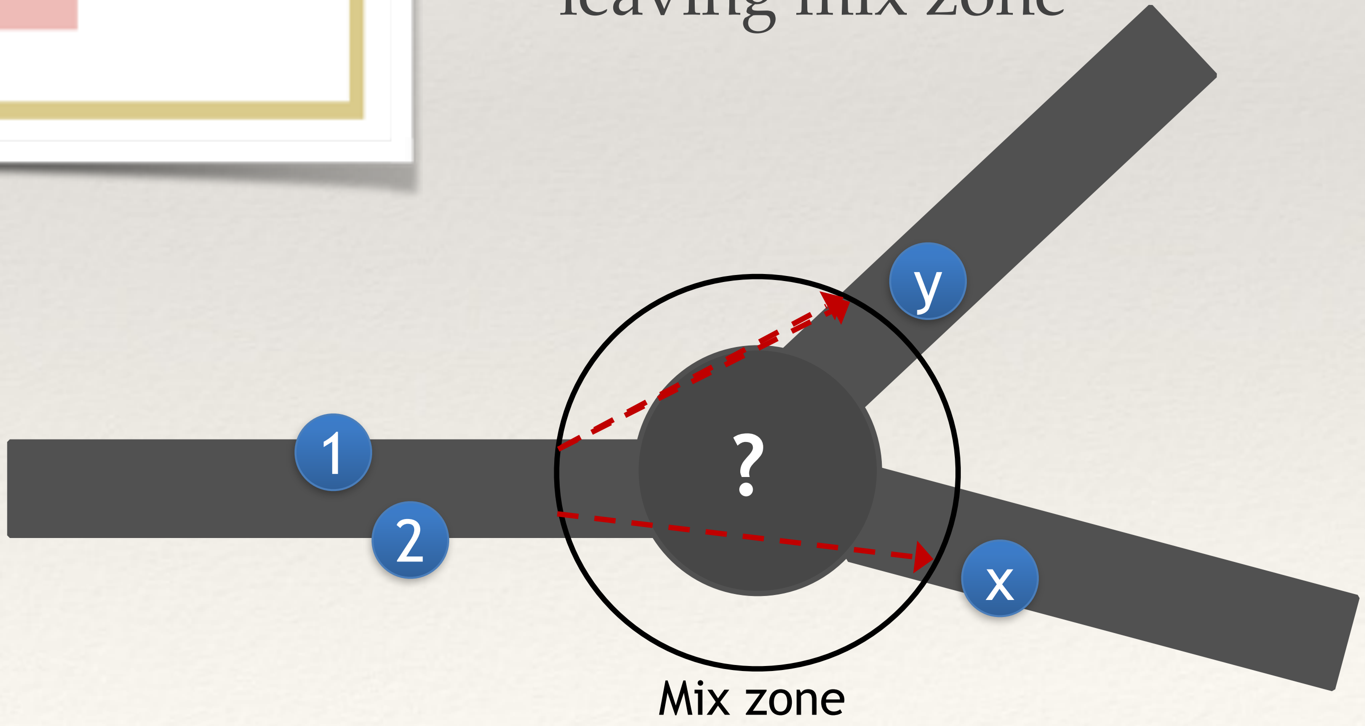
- Anonymity is measured as the fraction of mis-identified (40) location traces

R. Shokri, G. Theodorakopoulos, G. Danezis, JP. Hubaux, and JY. Le Boudec. "Quantifying Location Privacy: The Case of Sporadic Location Exposure", in PETS 2011

Defense: Mix Zone



- Threat: local eavesdroppers
- Spatial de-correlation: remain silent in mix zone
- Temporal de-correlation: change pseudonym after leaving mix zone

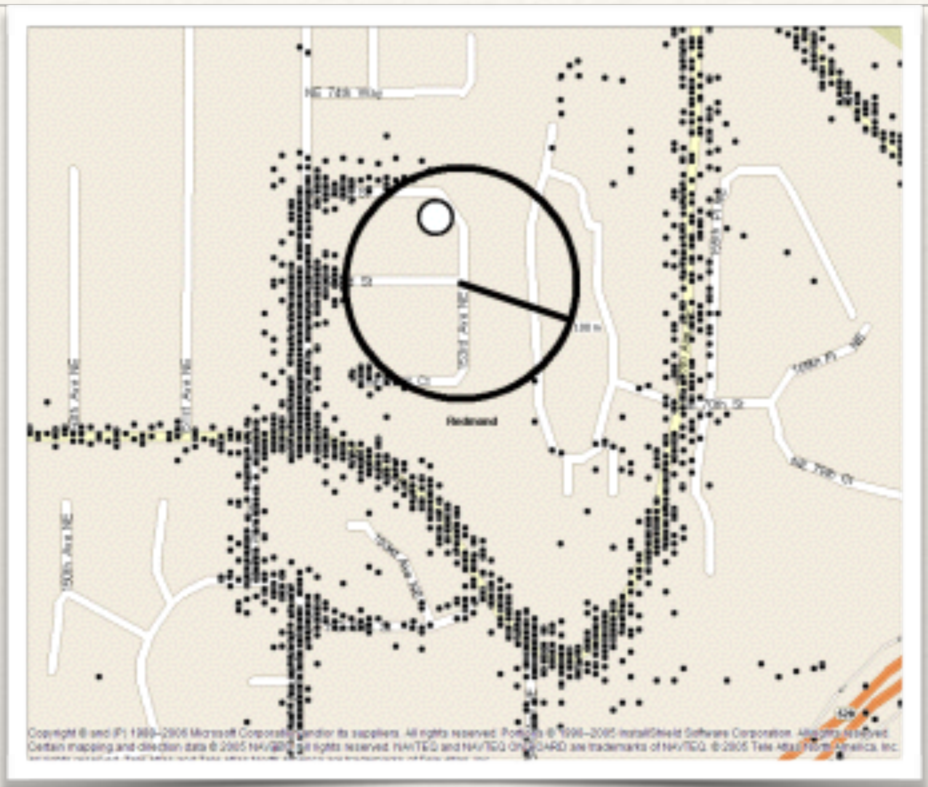


- Metric: Anonymity Set
- Limitations?

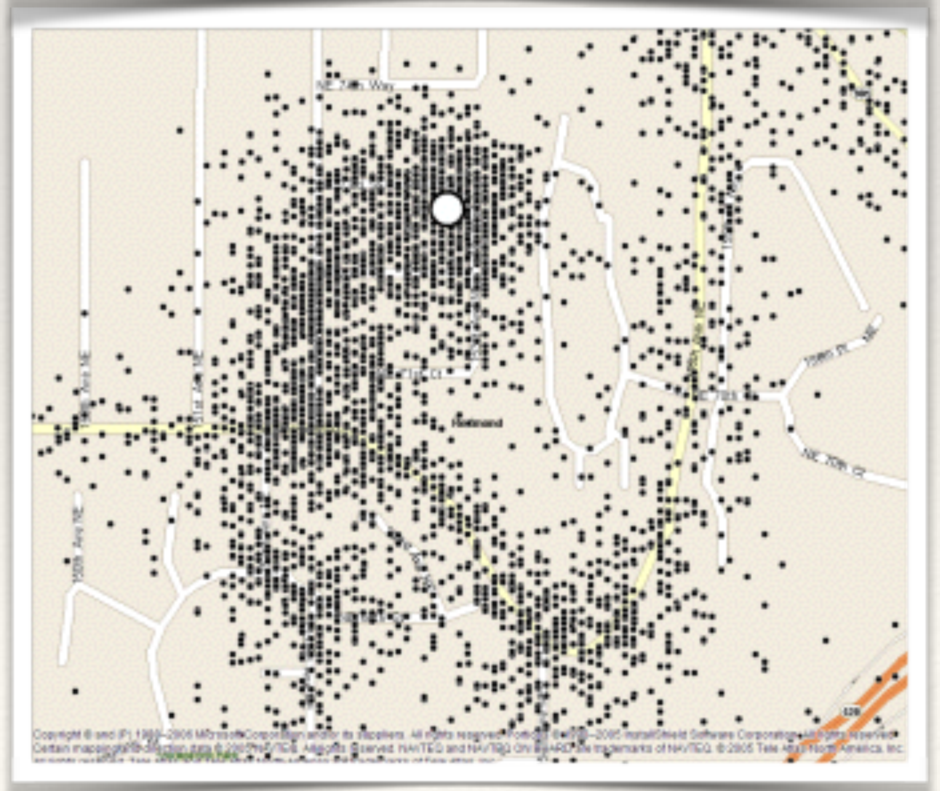
A. Beresford and F. Stajano. “Mix Zones: user privacy in location aware services”. In Percom, 2004

Defense: Obfuscate the Location

- Add noise to location information before sharing
- Cost? **Utility** loss?
- Limitations?
- Attack Resilience?



delete around Home



reduce accuracy

Defense: Path Confusion

- Anonymize all the location samples (remove likability between locations in a trace)
- Add noise to some locations to confuse a multi-target tracking (MTT) algorithms (that try to reconstruct original traces)
- Limitation? Complexity? Cost?
    - what if the confused traces are not geographically separate or belong to closely related people?

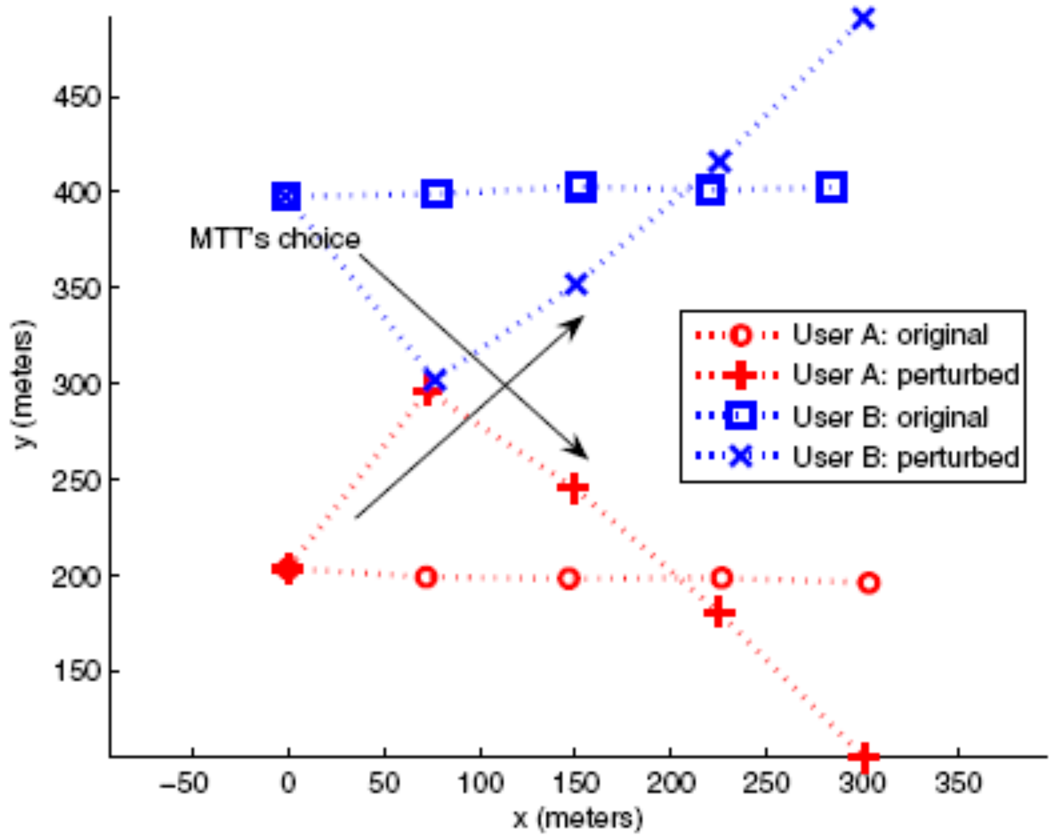
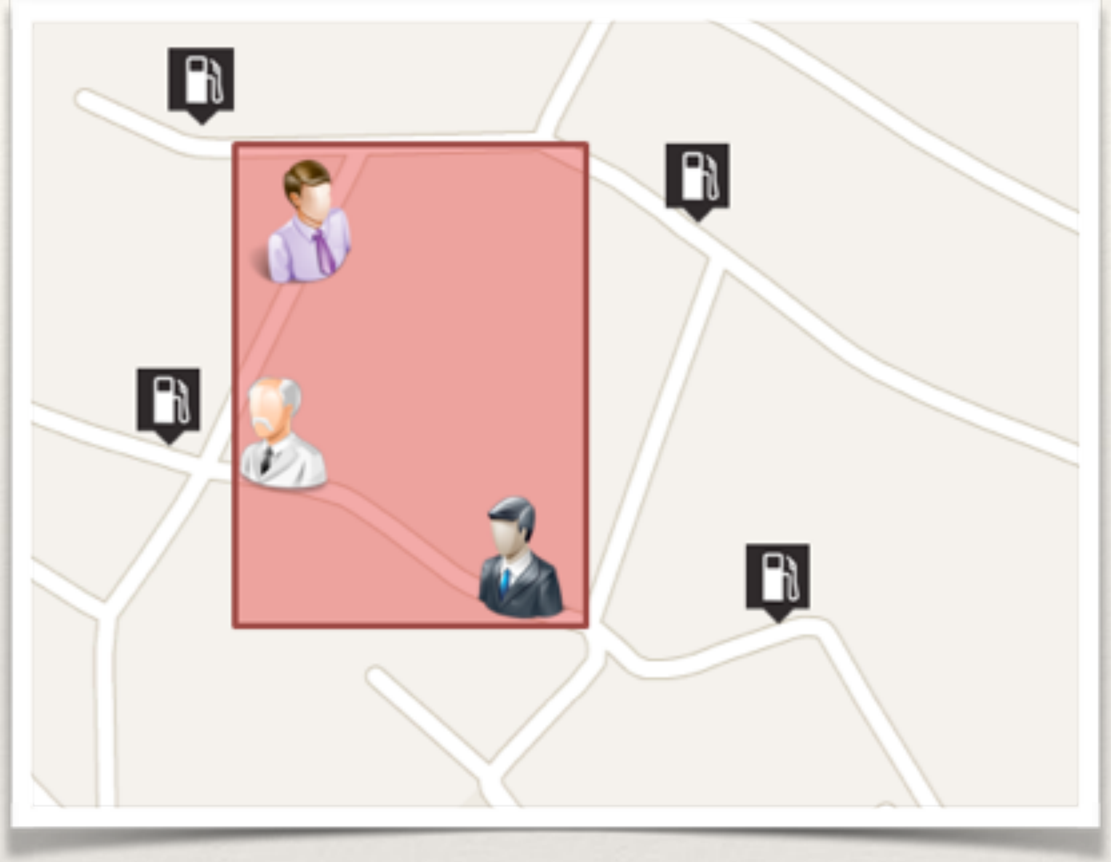


Figure 2. Two users move in parallel. The Path Perturbation algorithm perturbs the parallel segment into a crossing segment.

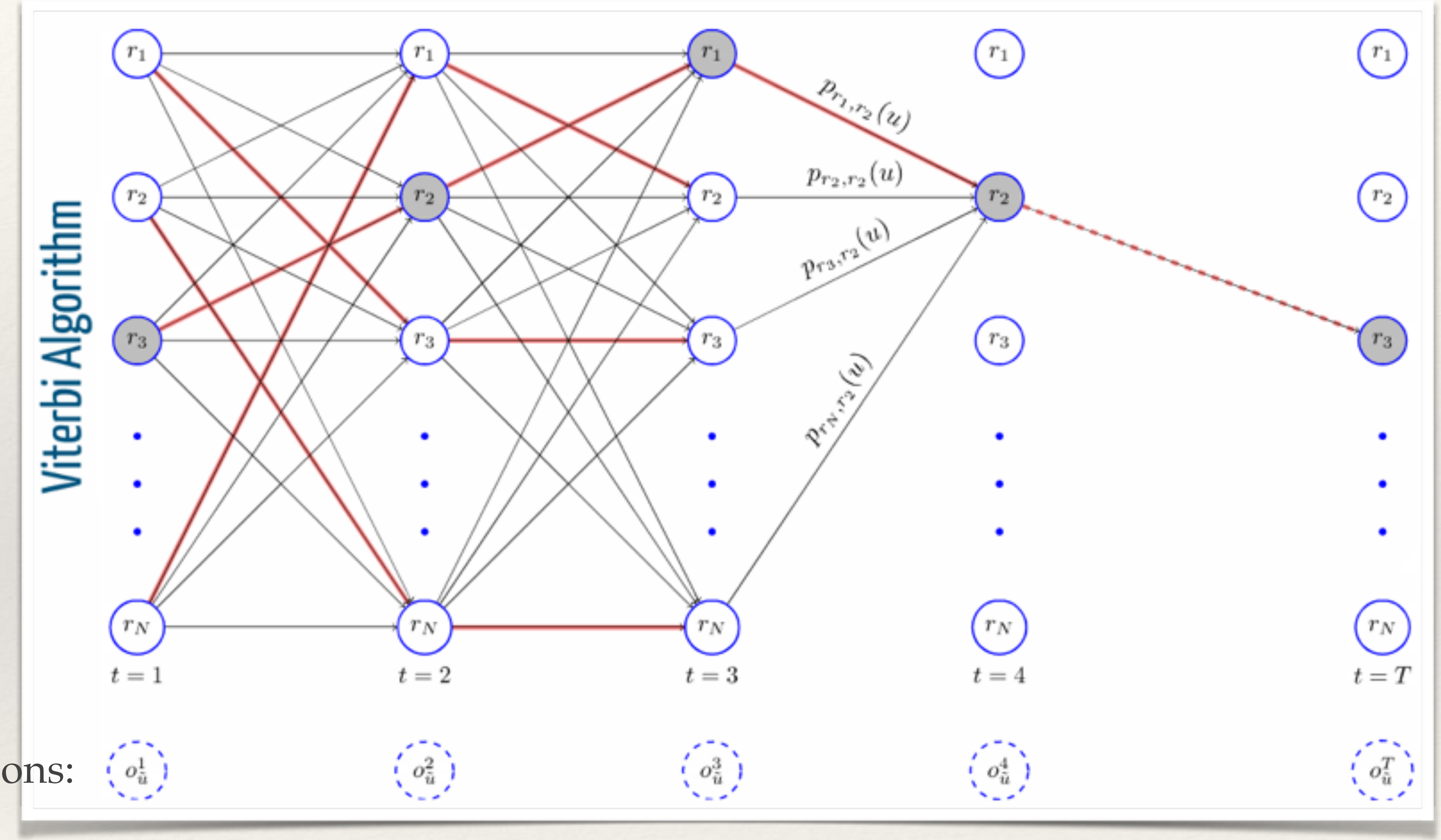
B. Hoh, M. Gruteser, “Protecting location privacy through path confusion”, In SECURECOMM 2005

Defense: Location K-Anonymity

- Location Cloaking: report a large area rather than your accurate location
- The cloaking area should be large enough such that it fully overlaps with that of k-1 other users
- Limitations?
- Attack resilience?



Attack: Tracking

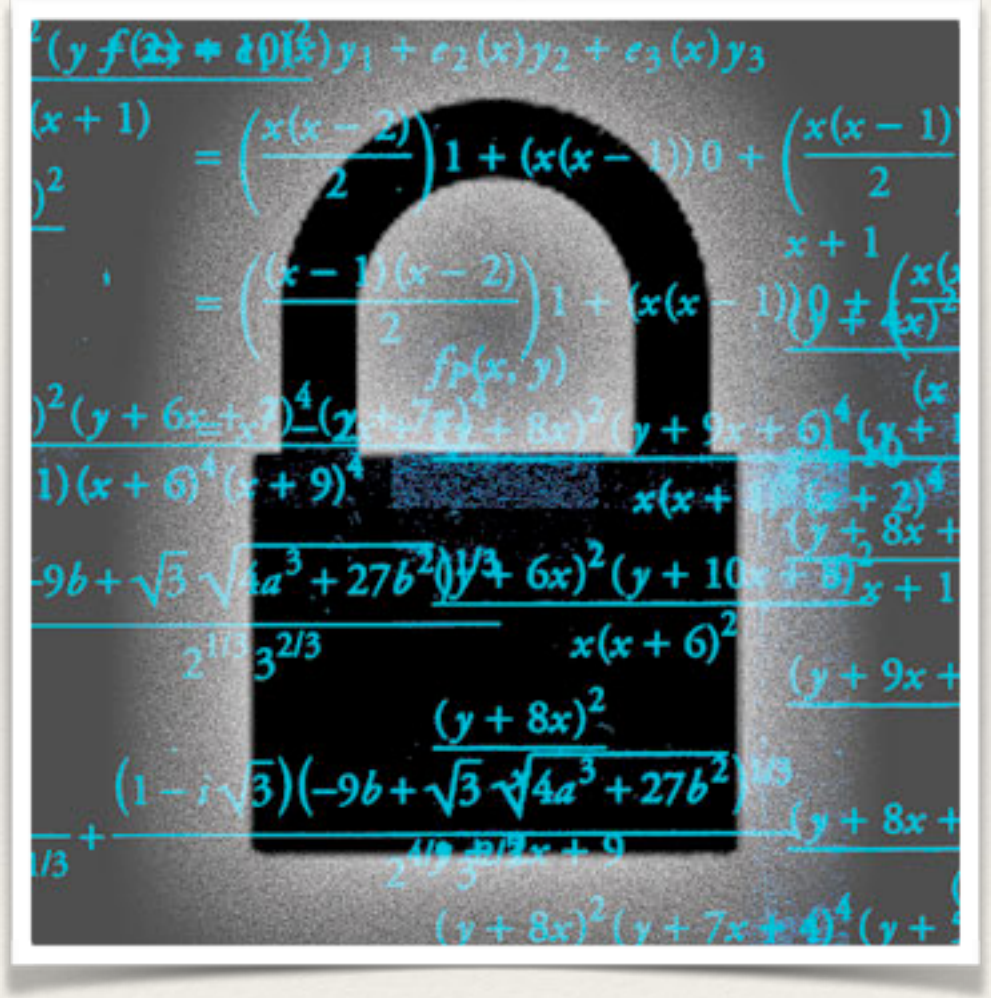


Observations:

- Each state is a different region/location where the user can visit
- Viterbi gives the most likely trace that could have produced a particular observation

Cryptographic Protocols

- Design a system that enables blind information processing (e.g., using homomorphic encryption)
- Cost?
  - computation complexity
- Limitation?
  - lack of service provider's incentive
- Attack Resilience?



Example: LBS using PIR

- Use private information retrieval (PIR) to obtain information about your whereabout when using a location-based service (LBS)
- LBS server has a database of contextual information about different locations.
- User specifies a search area and searches about points of interests around her location
- PIR enables searching and accessing information in a database without leaking information about the query to the database server

Privacy-wise it is equivalent to the case of downloading all the database associated with the (yellow) search area

F. Olumofin, P. K. Tysowski, I. Goldberg, U. Hengartner, "Achieving Efficient Query Privacy for Location Based Services". In 10th Privacy Enhancing Technologies Symposium, 2010.

Geo-Indistinguishability

- Add planar Laplacian noise to the location before sharing
- It satisfies “differential-privacy” for location data
- It guarantees that what an adversary knows after an observation is very close to what he knows prior to the attack

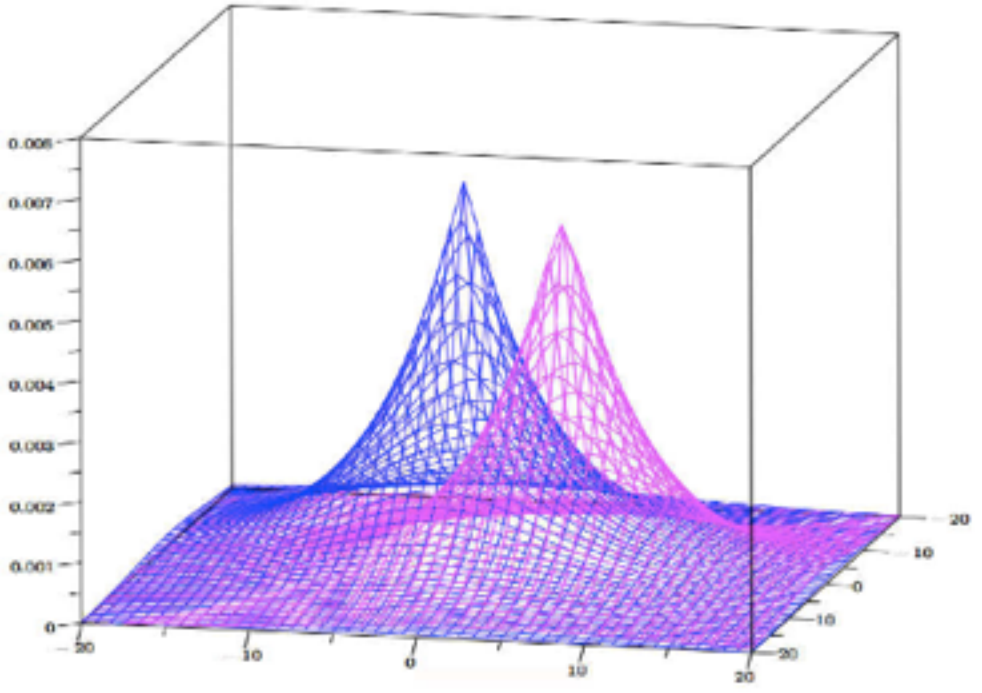
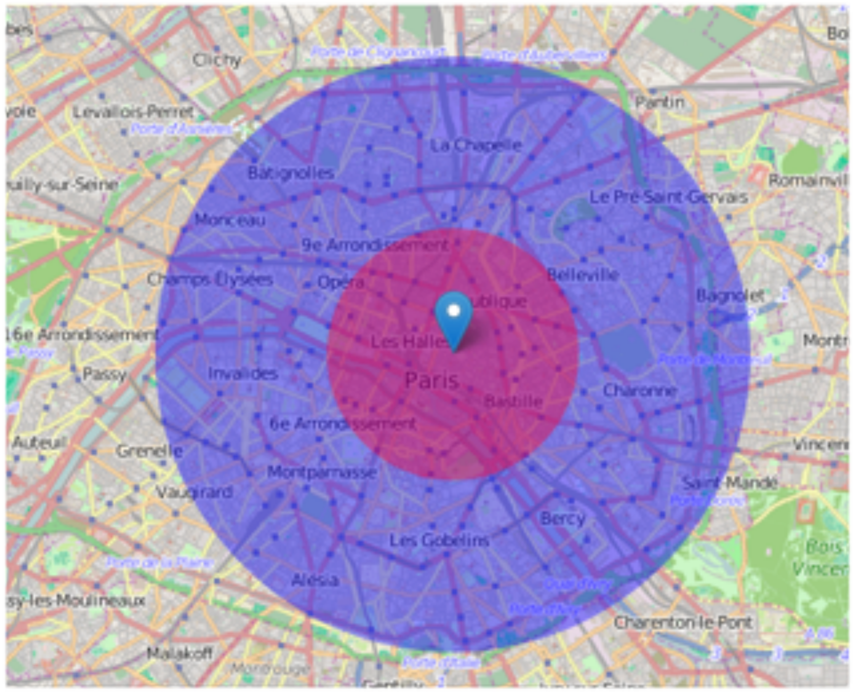
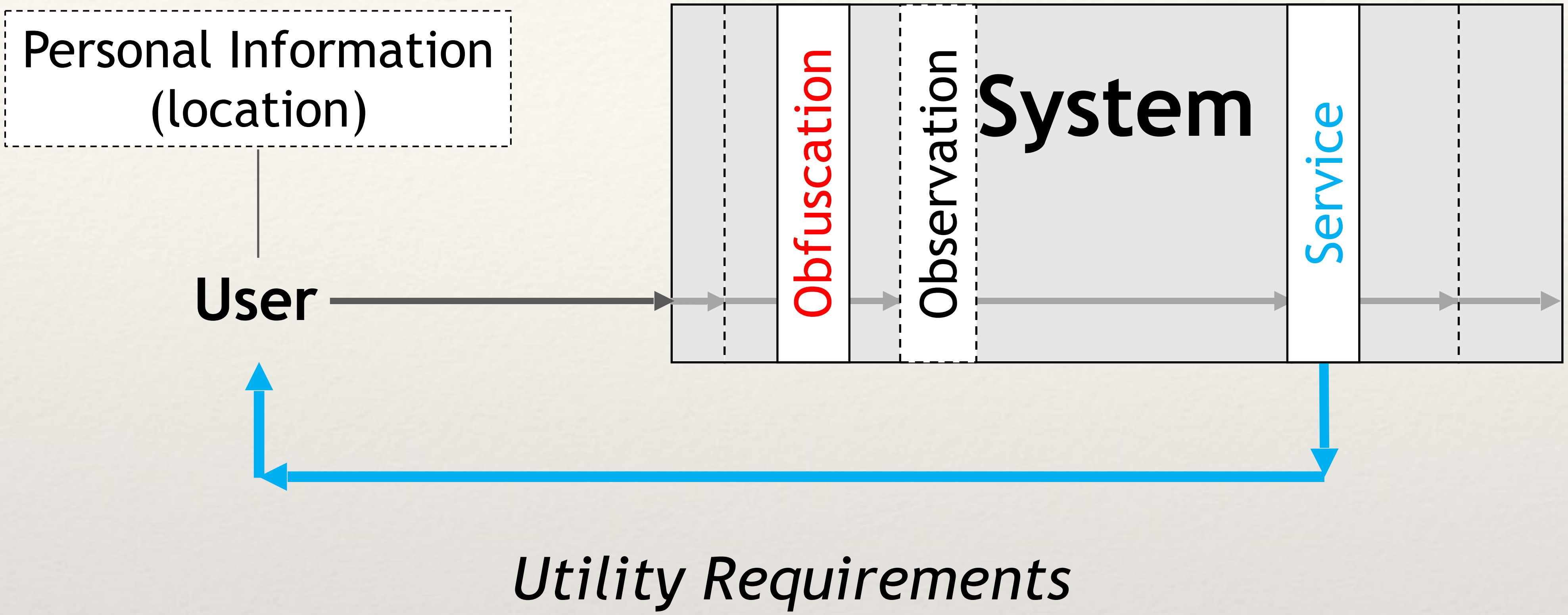


Figure 2: The pdf of two planar Laplacians, centered at $(-2, -4)$ and at $(5, 3)$ respectively, with $\epsilon = 1/5$.



M. E. Andres, et al., "Geo-Indistinguishability: Differential Privacy for Location-Based Systems", in CCS 2014

Optimal Obfuscation



- There is a tradeoff between privacy and utility
- Problem is to design an obfuscation mechanism that guarantees a minimum utility and maximizes location privacy
- We need metrics for both location privacy and utility

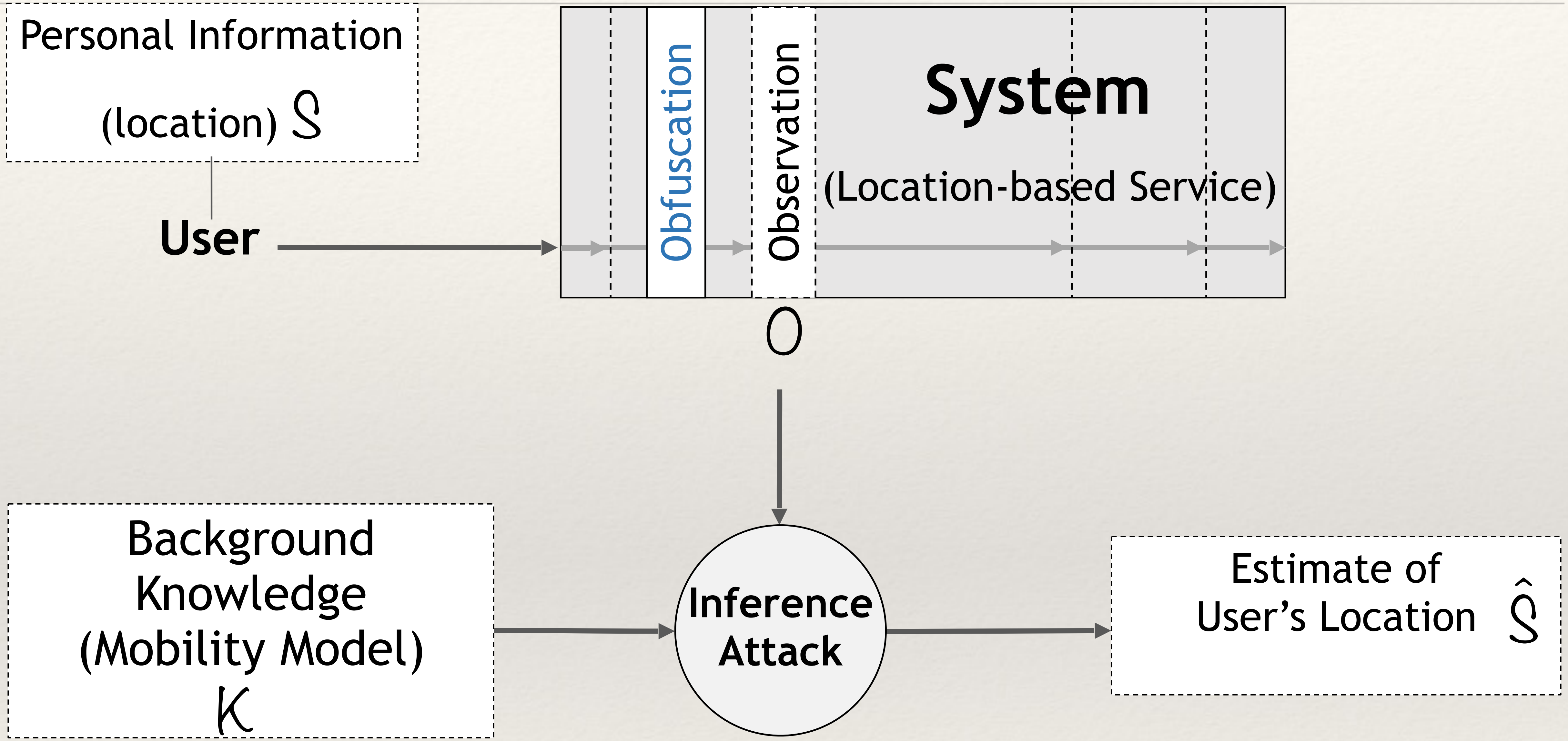
Quantifying Location Privacy

Metric

- Estimation Error: The error in correctly guessing someone's true location (at a given time, or during a time window)
- Background Knowledge: What is already known about the target (e.g., her name and work address)

R. Shokri, G. Theodorakopoulos, JY. Le Boudec, JP. Hubaux. "Quantifying Location Privacy", In IEEE Symposium on Security and Privacy, 2011.

Quantification Framework



Privacy (as expected inference error): $\sum_{\hat{S}} \Pr(\hat{S} \mid O,K) \, . \, d(S,\hat{S})$

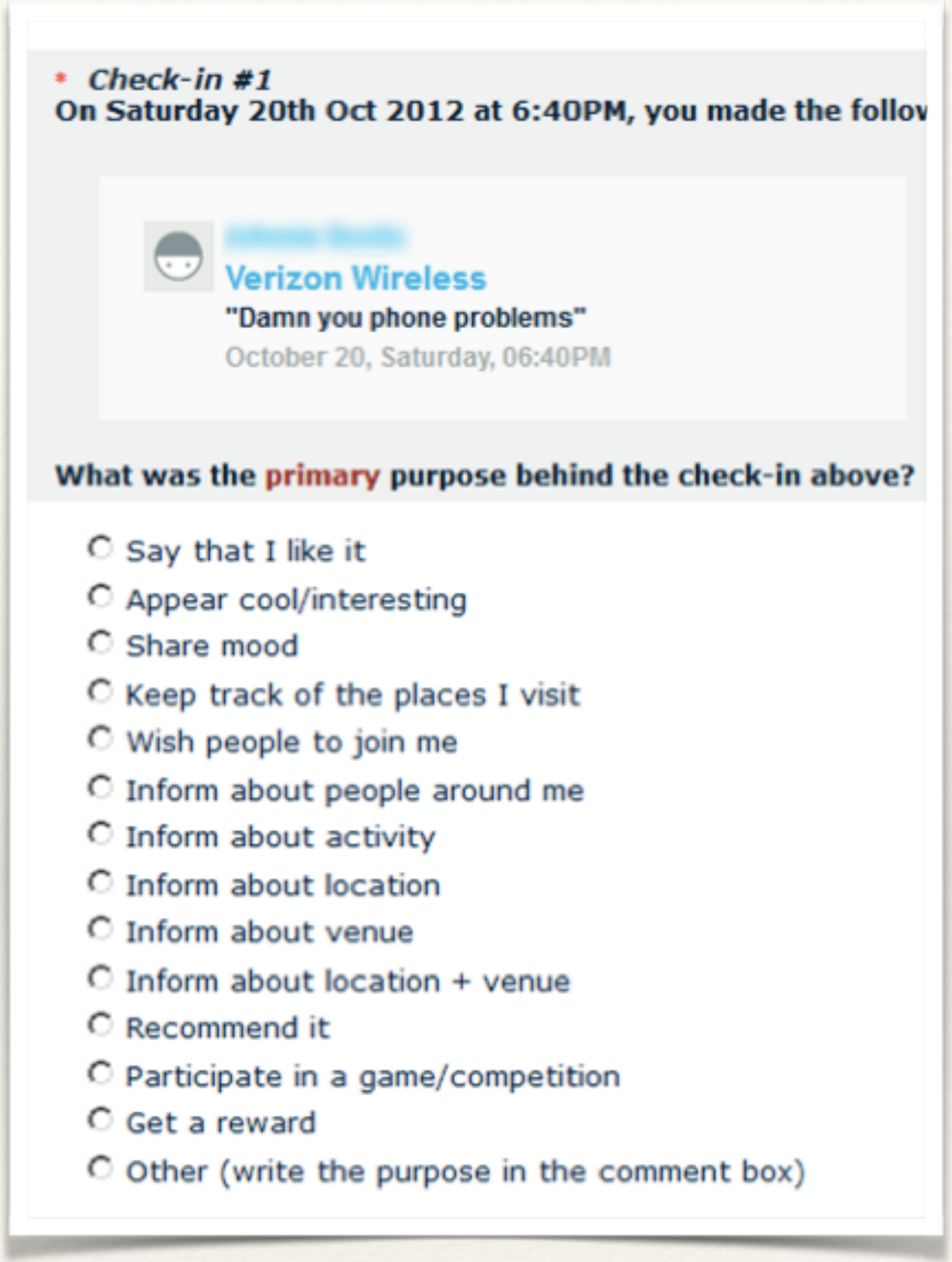
Cost of location privacy

Utility

is highly dependent on the motivation behind sharing a location

Survey people and ask them about the purpose of their location check-ins and to what extent they would be happy if an obfuscation is in place

Learn a function for utility using machine learning techniques



* ***Check-in #1***
On Saturday 20th Oct 2012 at 6:40PM, you made the follov

[redacted]
Verizon Wireless
"Damn you phone problems"
October 20, Saturday, 06:40PM

What was the primary purpose behind the check-in above?

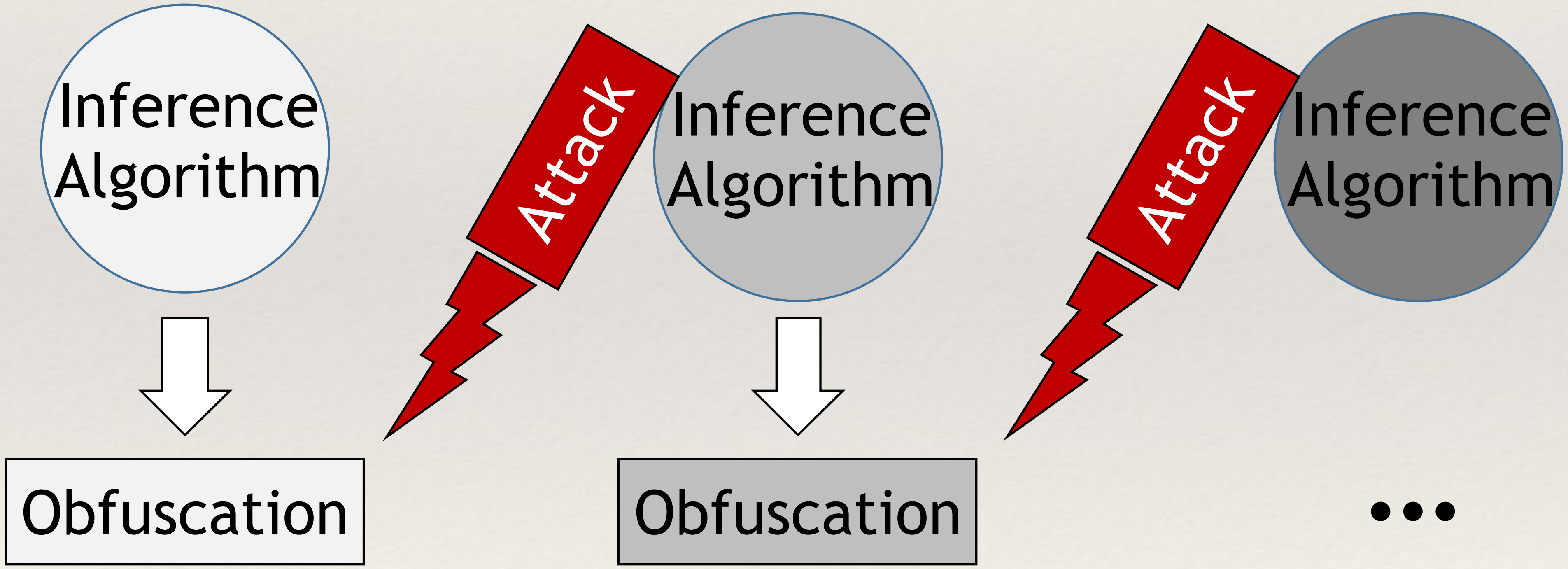
○ Say that I like it
○ Appear cool/interesting
○ Share mood
○ Keep track of the places I visit
○ Wish people to join me
○ Inform about people around me
○ Inform about activity
○ Inform about location
○ Inform about venue
○ Inform about location + venue
○ Recommend it
○ Participate in a game/competition
○ Get a reward
○ Other (write the purpose in the comment box)

I. Bilogrevic, et al. "Predicting Users' Motivations behind Location Check-Ins and Utility Implications of Privacy Protection Mechanisms", in NDSS 2015

Optimal Obfuscation

Solution: Decision Theory ?

- Minimize privacy loss
    - Satisfy utility constraints

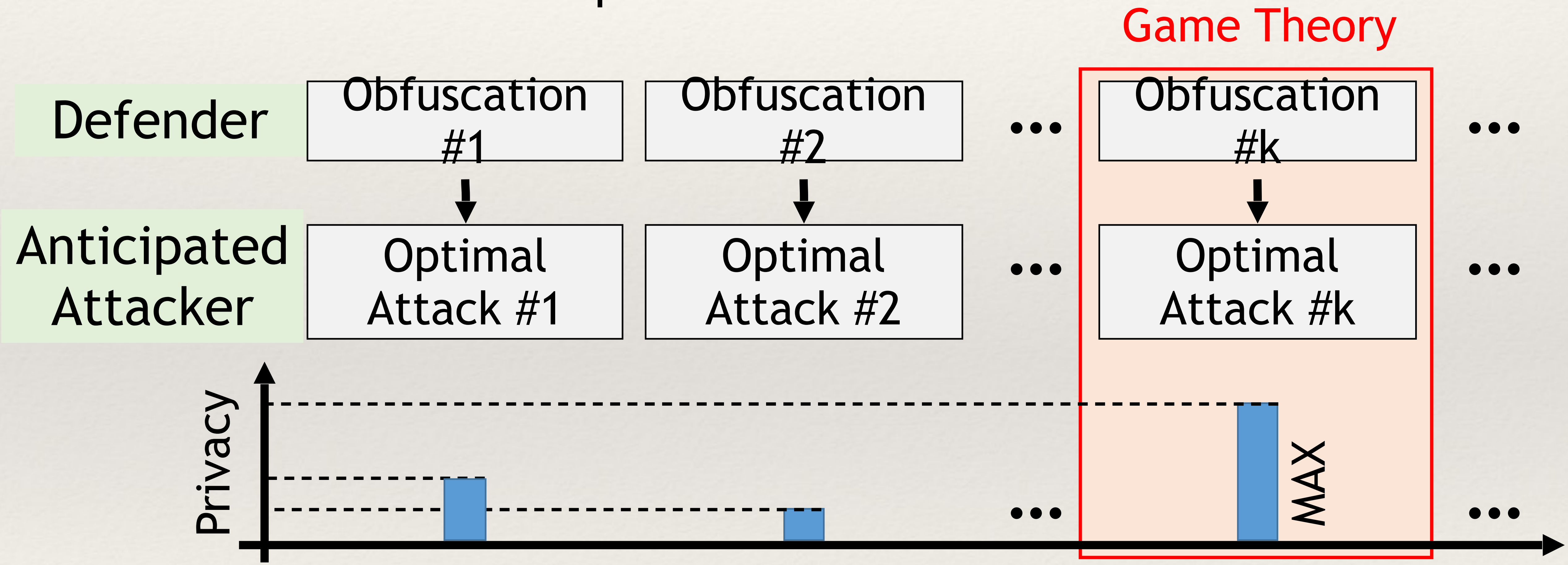


Privacy decision making must be interactive

Privacy Game

Attacker Has the Upper Hand

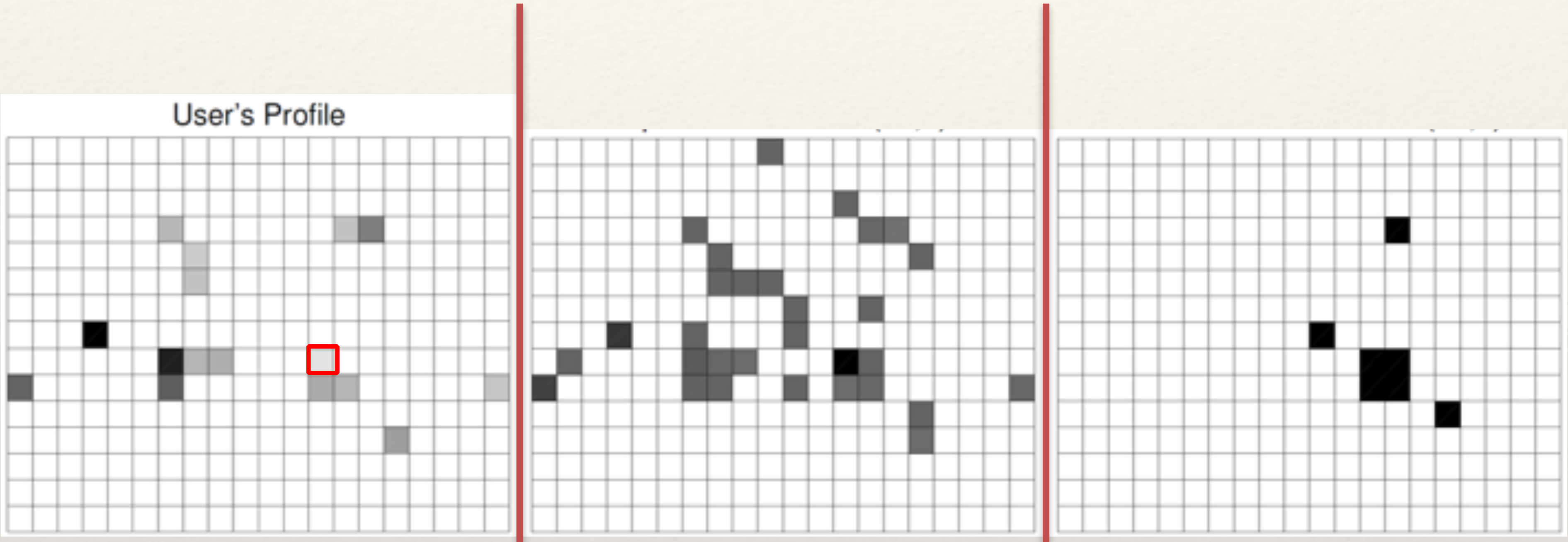
Defender Must Anticipate the Inference Attack



- Solve **conflicting** optimizations: Defense and Attack

R. Shokri, et al., "Protecting Location Privacy: Optimal Strategy against Localization Attacks," in ACM CCS 2012.

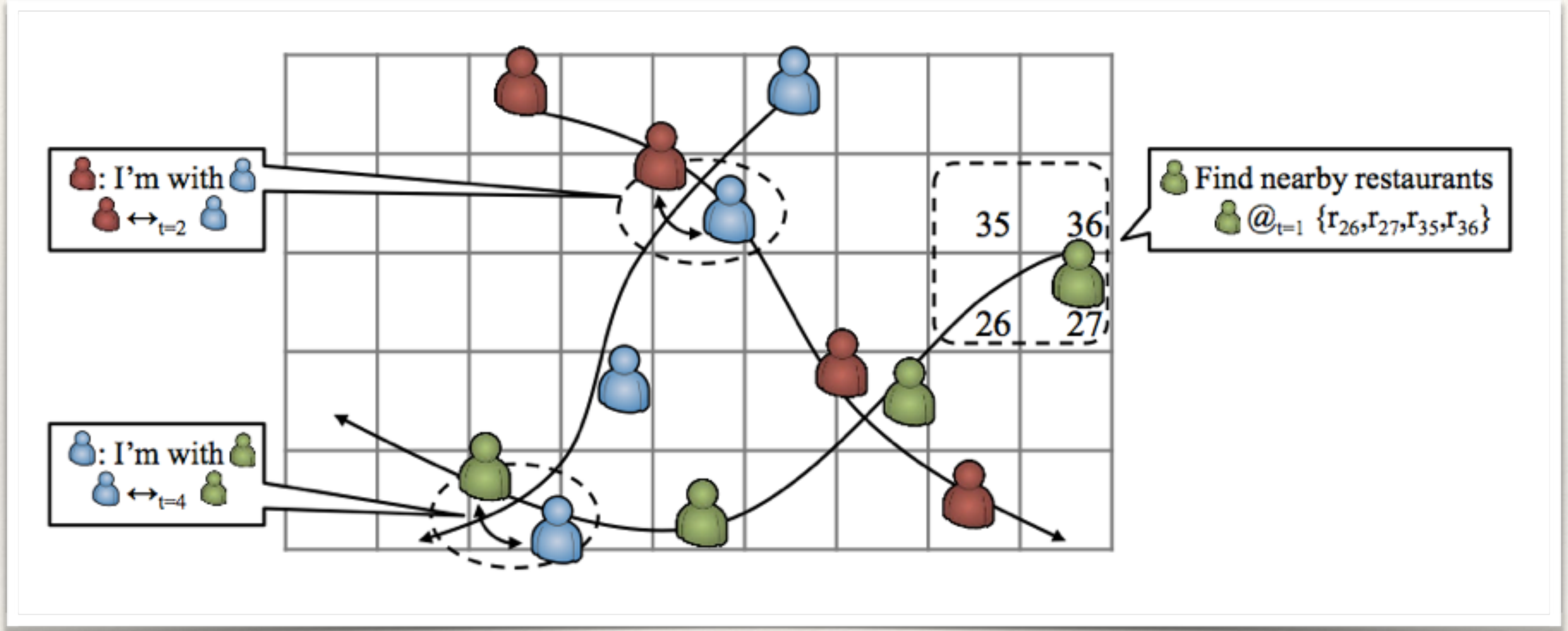
Output Visualization of Location Obfuscation



Optimal Obfuscation

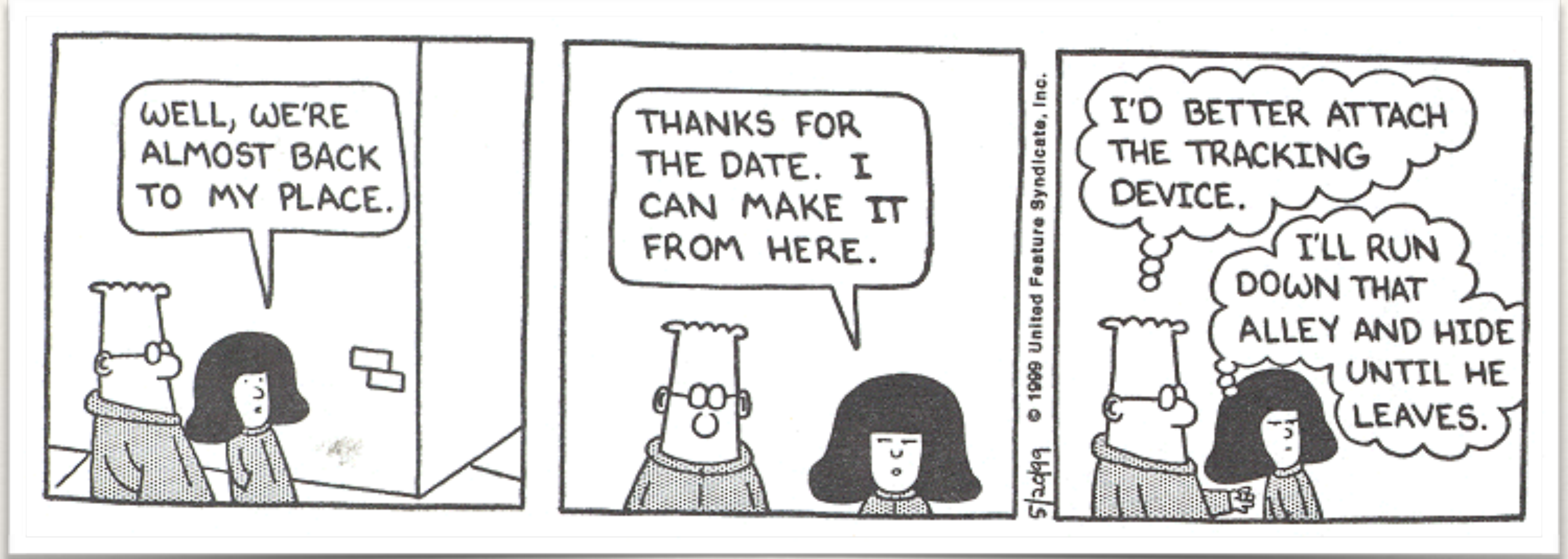
Basic Obfuscation (k = 7)

SociaLocation Privacy



- Social network can be inferred from location traces (e.g., NSA co-traveler program)
- Social co-location information can help an adversary to track users more accurately

A. M. Olteanu, et al. "Quantifying the Effect of Co-location Information on Location Privacy", in PETS'14


WELL, WE'RE ALMOST BACK TO MY PLACE.
THANKS FOR THE DATE. I CAN MAKE IT FROM HERE.
© 1999 United Feature Syndicate, Inc.
5/20/99
I'D BETTER ATTACH THE TRACKING DEVICE.
I'LL RUN DOWN THAT ALLEY AND HIDE UNTIL HE LEAVES.

NYC, Feb 18, 2015