Location Based Advertising and Location *k*-Anonymity

How can our location information be kept safe?

Matthew Gaba

Protecting Location Privacy with Personalized k-Anonymity: Architecture and Algorithms

Bugra Gedik, Ling Liu

http://scholar.google.com/scholar_url?hl=en&q=http://citeseerx.ist.psu.edu/viewdoc/download%3Fdoi%3D10.1.1.137.4420%26rep%3Drep1%26type%3Dpdf&sa=X&scisig=AAGBfm1lnG7y2j-ASa3_cdR0iAkVECQ4zQ&oi=scholarr

What is LBA?

- A form of advertising that uses user location data to serve users context-specific targeted advertisements.
- “Location data” can be collected from the user with or without the user’s knowledge and consent:
  - IP Geolocation
  - GPS
  - Cellphone Tower Triangulation
- Also may be supplied by the user:
  - Zip code
  - Address
  - Area code

Why is Location Data Valuable?

- Advertisers are paying almost 4x more for ad spots with location data, but why?
  - Demographic information
    - A rich or poor neighborhood?
    - Young or old?
    - Yale or Harvard?
    - Urban, suburban or rural?
  - Lifestyle choices
    - Safeway can “steal” costumers who frequent Target stores.
    - Did you run a 5k this weekend? Drink Gatorade!
  - Usage context
    - I am playing Angry Birds at home
      - Download Fruit Ninja Pro Advanced 3x!
    - I am playing Angry Birds in line at the supermarket
      - Use this coupon for Tide Laundry Detergent!

Location Data as Sensitive Information

- People are rightfully concerned about their location being tracked.
  - Controversy over iPhone logging information data.
- Just like location data reveals information to advertisers, it can reveal information to an adversary.
  - Political affiliation
  - Alternative Lifestyles
  - Medical Problems
  - Business practices

How do we protect this sensitive info?

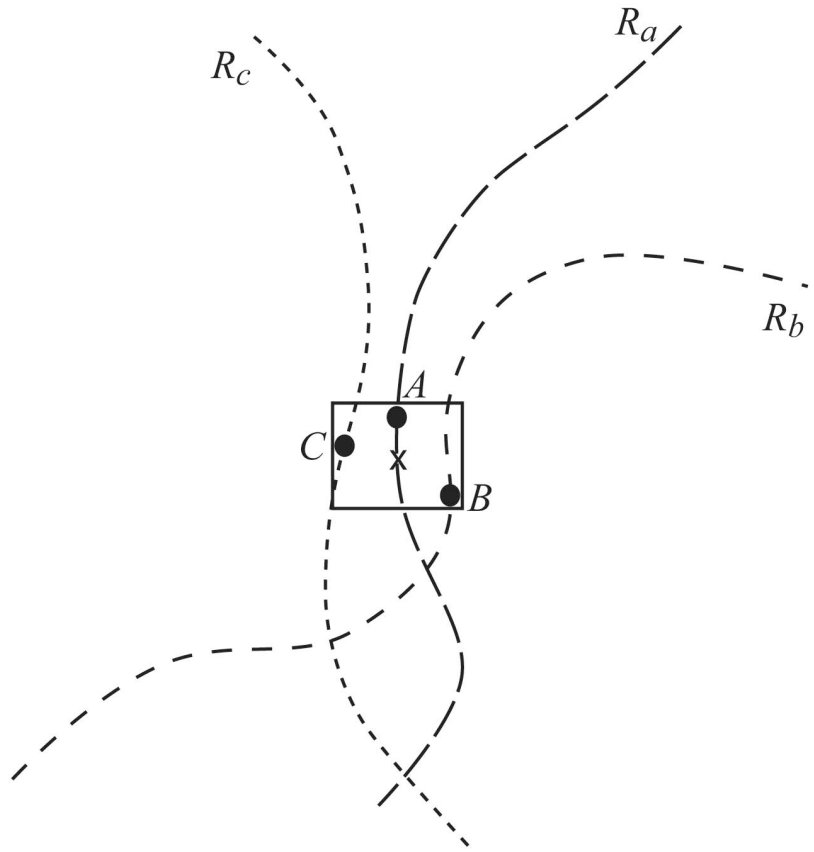
- We want to preserve the value of location data, while, at the same time, mitigating privacy risks.
  - This is tricky, because location data is inherently personal and private, and LBAs target individuals using this personal, private information.
- Can we parameterize this value-privacy tradeoff?
- Personalized *k*-Anonymity!

Personalized *k*-Anonymity

- A particular user's location is indistinguishable from *k*-1 other user's location.
- This protocol allows a user to choose their own *k* on a message-by-message basis.
- Also allows user to specify maximum acceptable loss-of-value of their personal information.

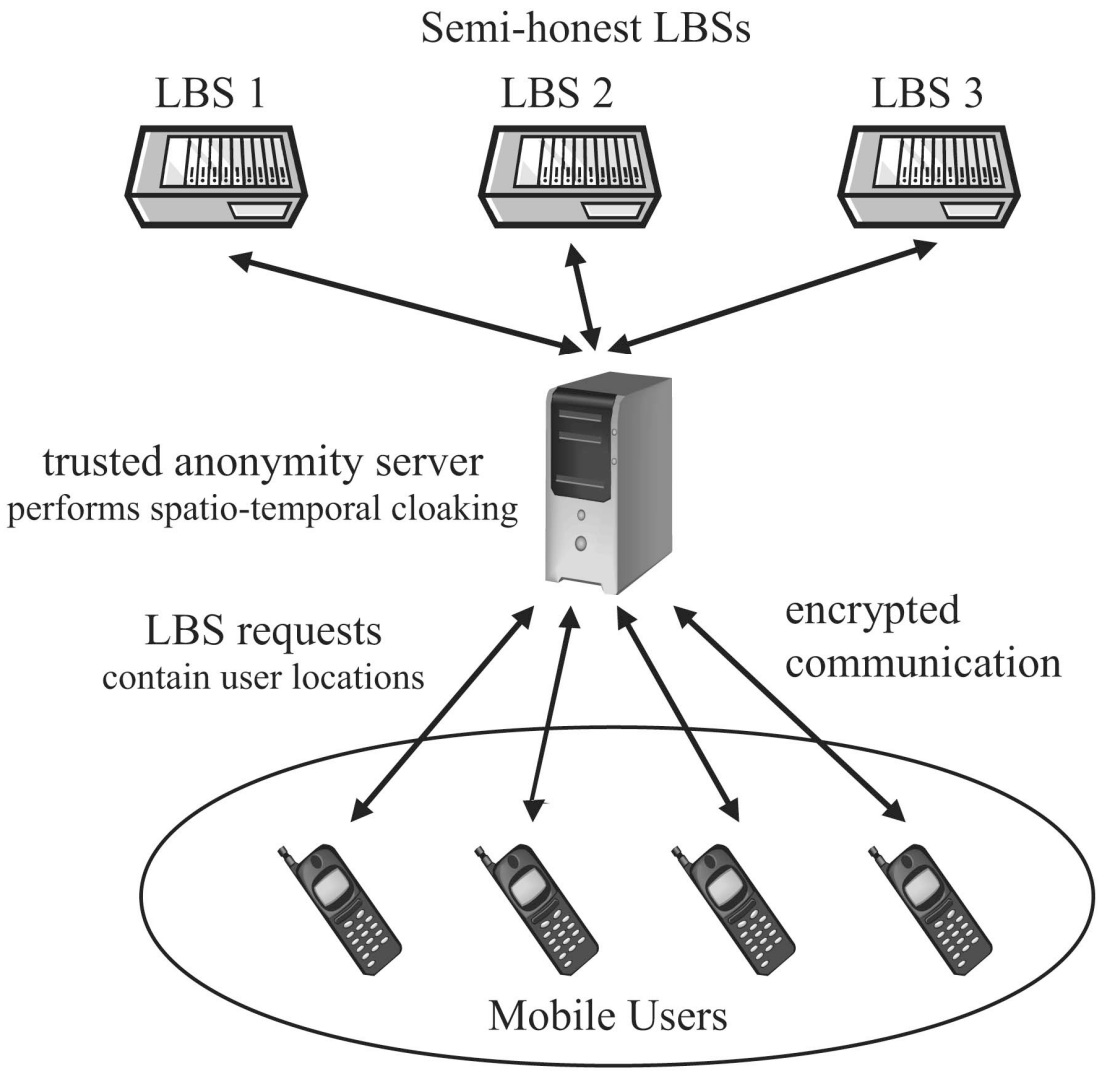
Why would a user want this?

- Why can't we just remove all PII from location messages? Anonymize each individual message.
- An adversary may still be able to identify an individual by using outside information.



Linking attack example.

Assumptions and Architecture



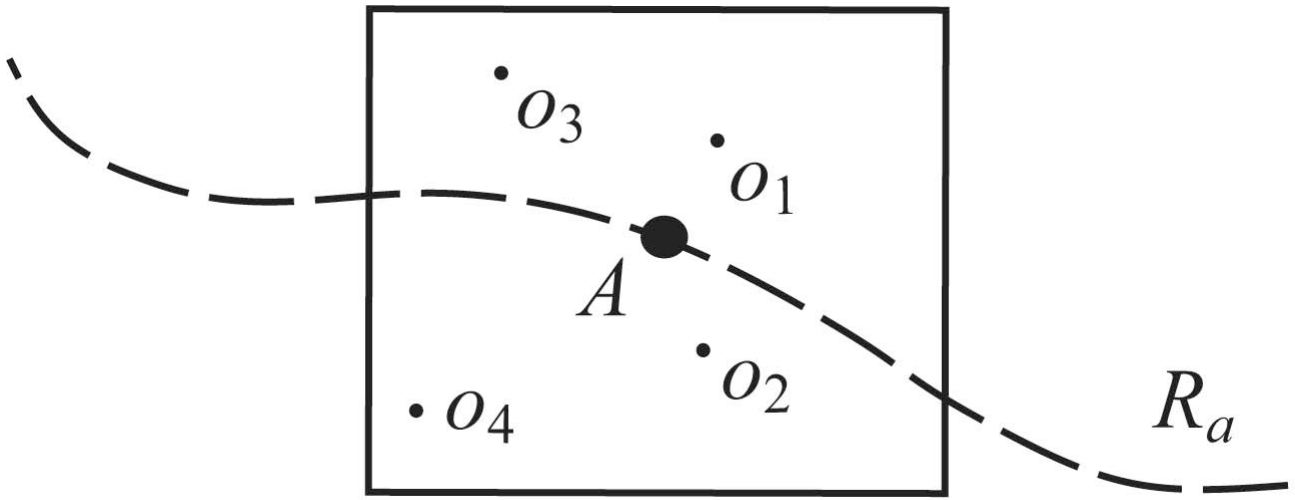
What's in a message?

$$m_s \in S : \langle \underbrace{u_{id}, r_{no}}_{\text{sender id, message no}}, \underbrace{\{t, x, y\}}_{\text{spatio-temporal point}}, \underbrace{k}_{\text{anonymity level}}, \underbrace{\{d_t, d_x, d_y\}}_{\text{temporal and spatial tolerances}}, C \rangle.$$

- Message sent to anonymizing server:
  - u_{id} : Sender Id
  - r_{no} : Message number
    - A message may be uniquely identified by u_{id} and r_{no}
  - t : Timestamp of message
  - x : x-coordinate of message
  - y : y-coordinate of message
    - Taken together, define a spatio-temporal point
  - k : Anonymity level
  - d_t, d_x, d_y : Temporal and spatial tolerance
  - C : Content of message

Need for Temporal and Spatial Tolerance

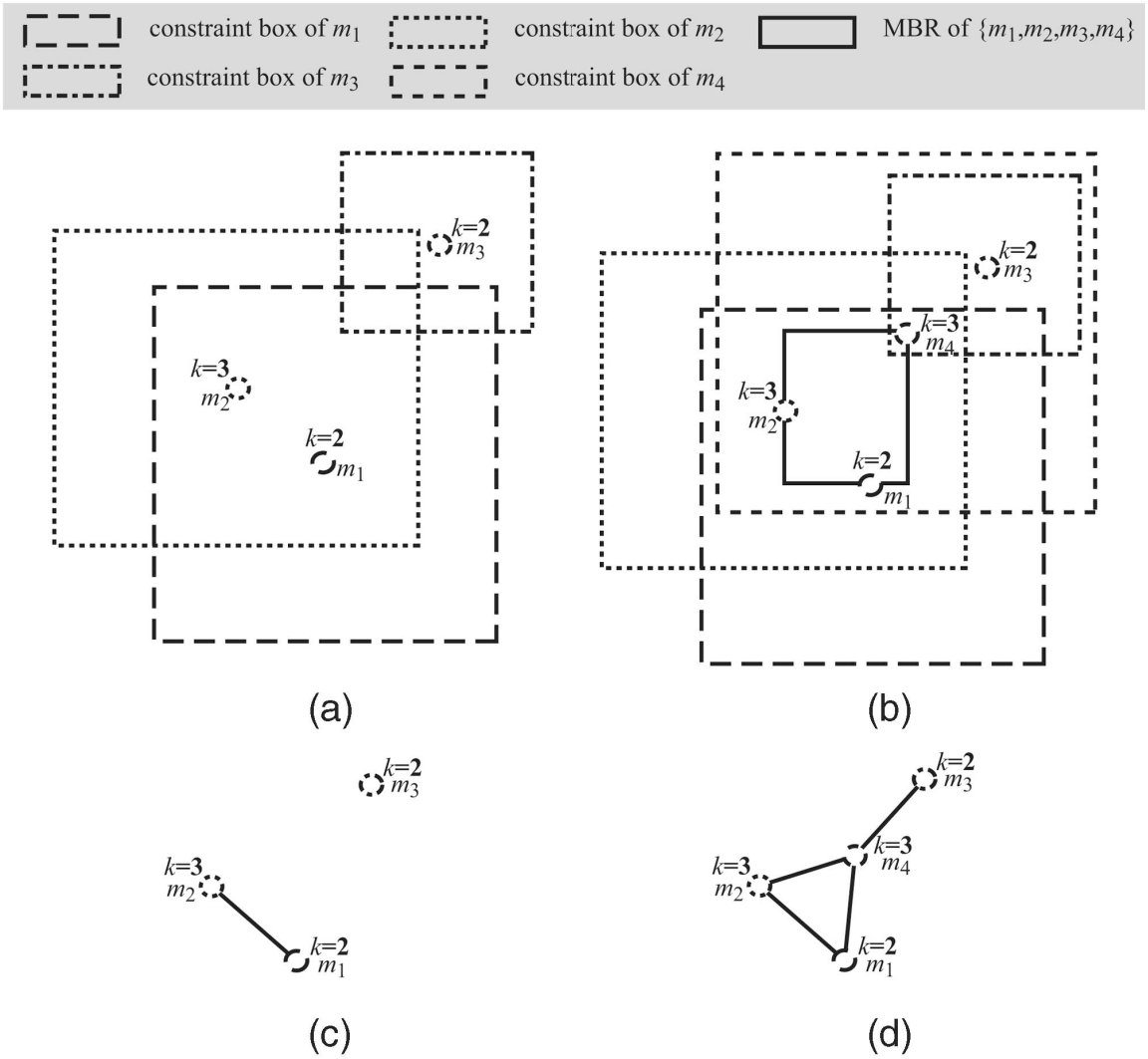
- Generally, achieving location *k*-anonymity with a higher k requires either a larger cloaking box, or longer temporal flexibility.
- Why is this bad?
- d_t, d_x, d_y components of message allow user to specify just how much loss of service (value) they are willing to tolerate.



The algorithm:

- Transform set of messages received by anonymity server into undirected graph.
- There exists an edge between two messages in the graph if and only if:
  - The messages originate from different mobile clients.
  - Their spatiotemporal points are contained in each other's constraint boxes defined by their tolerance values.
- Search graph for a clique s.t. size of clique is ≥ the max *k* value of all nodes in the clique.
  - Gedik and Liu give *CliqueCloak* algorithm for efficiently performing this operation.
- Compute smallest bounding box that contains all nodes in the clique.
- Server forwards this bounding box and a set of user identifiers corresponding to nodes in the bounding box to LBS providers.

Algorithm Illustration



Tradeoffs

Pros	Cons
■ Protects against several common types of attacks. ■ Allows a user to feel more secure in giving up sensitive location data. ■ Parameterization gives user control over their privacy.	■ Requires a trusted third party. ■ Not sure how this protocol would extend to the realm of advertising.

How can we use this in the realm of advertisements?

- A company like Google may want to give users the ability to adjust their *k* value in privacy settings
  - Price advertisers pay can scale with the size of the bounding box.
  - Google may be able to specify the temporal-spatial tolerance parameters if there is some cutoff point past which location data is meaningless to advertisers.
- A service provider, like Spotify, will offer a hybrid pay/ advertisement system. The service will allow a user to choose a *k*, and the higher the *k*-value, the more the person has to pay.
  - A clever way of fixing the privacy vs. pay problem of ad-supported services.
  - Complicated.

Questions?

Thoughts?