CS573 Data Privacy and Security

Location Privacy

Yonghui (Yohu) Xiao
http://yxiao.info
Outline

• What is Location Privacy

• Basic Techniques
  – Private Information Retrieval
  – Probabilistic Approach
    • Stationary
    • Temporal
What is Location Privacy

• Location Based Services (LBS)
  – Yelp, Google+, facebook, Instagram, Twitter...
  – Restaurant check-in, finding the nearest gas station, navigation, tourist city guide, ...

• Location Sharing
  – Find Friends, Find my iphone, ...

• Location Based Social Networks
  – Foursquare, Swarm

• Risks?
  – Give your location data for the service
Location Privacy

• Risks
  – Give your location traces to Google, Apple or other service providers
  – Enable malicious apps to know your locations
  – Sharing in Facebook, but available throughout internet
  – Locations may be leaked to other attackers through network
  – Physical danger, e.g. http://pleaserobme.com/

• User’s choices
  – Use LBS, give up privacy
  – Or preserve privacy, give up the LBS
  – Can we achieve the two goals: utility and privacy?
Features of Location Privacy

• Vs. Standard Differential Privacy
  – Differential Privacy: the outputs are similar whether a user opts in or out
  – For LBS, only one user

• Data Type
  – Standard Differential Privacy: tuples in Database
  – Location Privacy: where a user is

• Location data is only two-dimensional
  – Or at most three-dimensional
Techniques

• Encryption-based Techniques
  – Private Information Retrieval

• Probabilistic Techniques
  – Location obfuscation, location cloaking
  – Location generalization

• Continuous Protection
  – Temporal correlations
Private Information Retrieval (PIR)

• Allow user to query database while hiding the identity of the data-items she is querying.
  – What is the nearest restaurant to me?
  – Send a query to the server
  – Get a restaurant from the server

• Computational PIR
  – Homomorphic encryption

• Intensive Computational Cost
Probabilistic Techniques

• Spatial Cloaking/Location Generalization
  – Instead of sending the exact location to the service providers, a user can send a “general area”.

![Map of Paris with generalized locations](image)
Location Obfuscation

• Location Obfuscation
  – Instead of sending the exact location to the service providers, a user can send a “noisy” location.
  – Essentially, similar to spatial cloaking.
    • With the “general area”, a point can be randomly chosen to represent the “noisy” location.
    • The posterior probability of the “noisy” location will be the same as the “general area”. Can you prove it?
Probabilistic Techniques

• Privacy Guarantee
  – Uniform distribution in a circle
  – Uniform distribution in a polygon
  – Laplace distribution
  – Other distributions: 2D Gaussian distribution

• The trade-off between utility and privacy
  – What is the expected distance between the noisy location and the real location?
  – How much extra information does the noisy location give to attackers?
  – Can you derive the above distance function and the privacy function?
Geo-indistinguishability

• Geo-indistinguishability
  – A “differentially private” cloaking method
  – Based on the 2D Laplace distribution
  – Randomly draw a point from the distribution
Geo-indistinguishability

• Definition
  – Pr(z|x)≤e^{\epsilon}\cdot Pr(z|x')
  – Where x and x' are any two locations in a circle with a radius r, z is the noisy location

• Features
  – Location data: x and x’ are two points on a map
  – Neighboring databases: any points in the circle
  – Protection: indistinguishability in the circle
Geo-indistinguishability

• Geo-indistinguishability
  – How to prove the privacy?

\[
\begin{align*}
D_{\epsilon}(x_0)(x) &= \frac{\epsilon^2}{2\pi} e^{-\epsilon d(x_0, x)} \\
\end{align*}
\]

  – How much differential privacy can it provide?

• Open question:
  – Can you come up with a better sampling algorithm than the paper (Geo-indistinguishability, CCS13)
Continuous Approach

• Potential problems of the cloaking algorithms at stationary timestamps.
  – Not private in a period of time.
  – Examples:
Continuous Approach

- Location Release over time
Continuous Approach

• Temporal Correlations
  – Road network
  – Moving patterns of a user
  – Example:
    • Given that Alice is at MSC building now, she may go to Starbucks with probability 0.3, DUC with probability 0.3, and library with probability 0.4.

• How to describe such correlations?
  – A common method is to use Markov model
Markov Model

- Markov Model
  - Coordinate System
Markov Model

- Markov Model
  - Transition Matrix
    - A matrix $M$ denotes the probabilities that a user moves from one location to another
    - $M_{i,j}$ is the probability of moving from location $i$ to location $j$.
      - $M_{i,j}$ is the element of $ith$ row and $jth$ column.
Markov Model

- Markov Model
  - Emission Probability
    - Given the real location $i$, what is the probability distribution of the noisy locations?
    \[
    Pr(z_t | u^*_t = s_i)
    \]
  - Inference and Evolution
    \[
    p^+_t[i] = Pr(x_t = s_i | z_t) = \frac{Pr(z_t | x_t = s_i)p^+_t[i]}{\sum_j Pr(z_t | x_t = s_j)p^+_t[j]}
    \]
Markov Model

• Derive the possible locations at current timestamp
  – Bayesian inference using the previously released locations.
  – A set of possible locations can be generated.

• Only protect the true location within this set of possible locations.
  – Recall the definition of “neighboring databases”
  – What is the new neighboring databases here?
Extended Differential Privacy

Definition (Differential Privacy)
At any timestamp $t$, a randomized mechanism $A$ satisfies $\epsilon$-differential privacy on $\delta$-location set if, for any output $z_t$ and any two locations $x_1$ and $x_2$ in $\delta$-location set, the following holds:

$$
\frac{Pr(A(x_1) = z_t)}{Pr(A(x_2) = z_t)} \leq e^\epsilon
$$

Intuition
the released location $z_t$ (observed by the adversary) will not help an adversary to differentiate any instances in $\delta$-location set.
Probability Design

• Design a distribution on the set of possible locations.
Continuous Released Locations

- Example: Released “Noisy” Locations
References

• Geo-indistinguishability: differential privacy for location-based systems, CCS, 2013
• Protecting Locations with Differential Privacy under Temporal Correlations, CCS, 2015
• Quantifying Location Privacy, IEEE SP 2011
• In-Network Trajectory Privacy Preservation, ACM Computing Surveys (CSUR) 2015