Steps Towards Location Privacy

Subhasish Mazumdar

New Mexico Institute of Mining & Technology
Socorro, NM 87801, USA.

DataSys 2018

Census

- A census is vital for a country's planned growth.
- Census data must result in public information.
- There is need to assure citizens that collected data would help them *without* revealing their personal information.
  Otherwise, collected data would lead to incomplete and/or inaccurate information.

Statistical Queries

- Each individual has *identifiable* attributes like name, *shared* attributes like gender, zip code, and *sensitive* attributes like income, political affiliation / contribution.
  Attack: Retrieve based on attributes knowing there is just one matching individual.

Statistical Queries

- Each individual has *identifiable* attributes like name, *shared* attributes like gender, zip code, and *sensitive* attributes like income, political affiliation / contribution.
  Attack: Retrieve based on attributes knowing there is just one matching individual.
- Solution: Disallow aggregates for *small* sets.

Statistical Queries

- Each individual has *identifiable* attributes like name, *shared* attributes like gender, zip code, and *sensitive* attributes like income, political affiliation / contribution.
  Attack: Retrieve based on attributes knowing there is just one matching individual.
- Solution: Disallow aggregates for *small* sets.
- Attack: Retrieve the complement of the small set.

Statistical Queries

- Each individual has *identifiable* attributes like name, *shared* attributes like gender, zip code, and *sensitive* attributes like income, political affiliation / contribution.
  Attack: Retrieve based on attributes knowing there is just one matching individual.
- Solution: Disallow aggregates for *small* sets.
- Attack: Retrieve the complement of the small set.
- Solution: disallow query results when either the result set *or its complement* is small.
  Allow results of size $[k, (N-k)]$ for $N \geq k > 1$

Statistical Queries

- Each individual has *identifiable* attributes like name, *shared* attributes like gender, zip code, and *sensitive* attributes like income, political affiliation / contribution.
  Attack: Retrieve based on attributes knowing there is just one matching individual.
- Solution: Disallow aggregates for *small* sets.
- Attack: Retrieve the complement of the small set.
- Solution: disallow query results when either the result set *or its complement* is small.
  Allow results of size $[k, (N-k)]$ for $N \geq k > 1$
- Attack: Schlörer, Palme.
  Attack: Denning and Denning: Idea: pad the small query sets with irrelevant records to make them large enough.
  Irrelevant records from a mask M.

[The Tracker: A Threat to Statistical Database Security. D.E. Denning, P.J. Denning, and M.D.Schwartz. ACM TODS 4(1) 1979.]

Example

R(ID, Gender, Dept, Position, Salary, Political contribution)

Alice is the only female CS professor.

Mask $M =$ males

Example

R(ID, Gender, Dept, Position, Salary, Political contribution)

Alice is the only female CS professor.

Mask M = males

1. count of employees in mask M
2. count of employees not in mask M (get N)

Example

R(ID, Gender, Dept, Position, Salary, Political contribution)

Alice is the only female CS professor.

Mask M = males

1. count of employees in mask M
2. count of employees not in mask M (get N)
3. sum of salary of employees in mask M
4. sum of salary of employees not in mask M (get total salary S)

Example

R(ID, Gender, Dept, Position, Salary, Political contribution)

Alice is the only female CS professor.

Mask M = males

1. count of employees in mask M
2. count of employees not in mask M (get N)
3. sum of salary of employees in mask M
4. sum of salary of employees not in mask M (get total salary S)
5. count of female CS professors or M
6. count of female CS professors or $\neg M$ ($-N$ to get 1 female CS prof)

Example

R(ID, Gender, Dept, Position, Salary, Political contribution)

Alice is the only female CS professor.
Mask M = males

1. count of employees in mask M
2. count of employees not in mask M (get N)
3. sum of salary of employees in mask M
4. sum of salary of employees not in mask M (get total salary S)
5. count of female CS professors or M
6. count of female CS professors or $\neg M$ ($-N$ to get 1 female CS prof)
7. sum of salary of female CS professors or M
8. sum of salary of female CS professors or $\neg M$ ($-S$ to get Alice's sal)

Medical Data

- Cities, hospitals, ... possess data about people that needs to be released.
- We want to release information about hospital visits and associated medical conditions but we do not want anyone to find out *who* had which condition.

Medical Data

- Cities, hospitals, ... possess data about people that needs to be released.
- We want to release information about hospital visits and associated medical conditions but we do not want anyone to find out *who* had which condition.
- Sharing medical data benefits society.
- Without *preserving anonymity*, they hurt the people they serve.
- Solution: Sanitize (eliminate identifying attributes) and release: Eliminate *explicit identifiers*, e.g., ID, name, address, phone number.

Sanitized released data

Massachusetts
Medical Data

Gender
DoB
Zip
Ethnicity
Visit Date
Diagnosis
Procedure
Medication
Total Charge

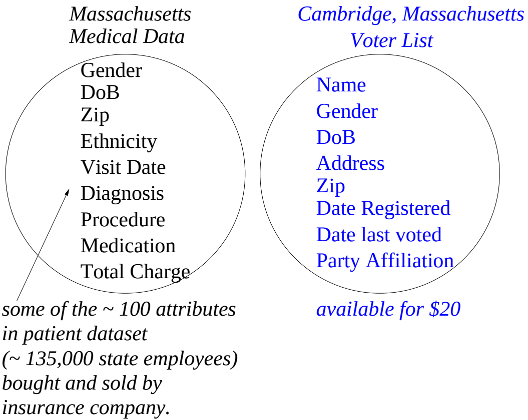
some of the ~ 100 attributes
in patient dataset
(~ 135,000 state employees)
bought and sold by
insurance company.

Attack: Re-identification through linking

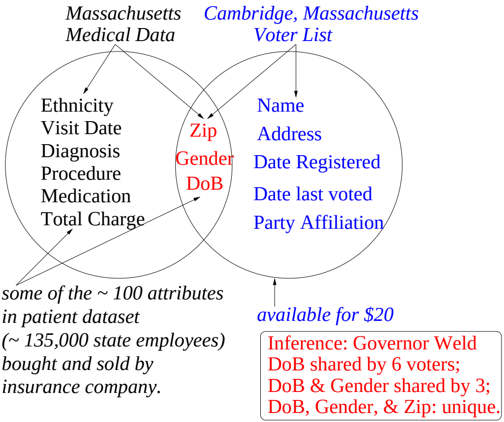
- Attack (Sweeney): *Re-identification* is possible by
  - linking (i.e., matching shared attributes) the released data with other available datasets ; or by
  - examining the distribution of the attributes.

[L. Sweeney. *k*-anonymity: a model for protecting privacy. Intl. J. on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002.]

Linking


Massachusetts
Medical Data
Gender
DoB
Zip
Ethnicity
Visit Date
Diagnosis
Procedure
Medication
Total Charge
some of the ~ 100 attributes
in patient dataset
(~ 135,000 state employees)
bought and sold by
insurance company.
Cambridge, Massachusetts
Voter List
Name
Gender
DoB
Address
Zip
Date Registered
Date last voted
Party Affiliation
available for $20

Linking



Massachusetts
Medical Data
Cambridge, Massachusetts
Voter List
Ethnicity
Visit Date
Diagnosis
Procedure
Medication
Total Charge
Zip
Gender
DoB
Name
Address
Date Registered
Date last voted
Party Affiliation
some of the ~ 100 attributes
in patient dataset
(~ 135,000 state employees)
bought and sold by
insurance company.
available for $20
Inference: Governor Weld
DoB shared by 6 voters;
DoB & Gender shared by 3;
DoB, Gender, & Zip: unique.

Voter list for Cambridge, MA

- Possible values: 2 genders; 5 zip codes (5-digit), 365×100 dob's $\Rightarrow$ 365,000 unique values are possible;
- Size was 54,805.

Voter list for Cambridge, MA

- Possible values: 2 genders; 5 zip codes (5-digit), 365×100 dob's $\Rightarrow$ 365,000 unique values are possible;
- Size was 54,805.
- *12%* had unique dob [month, day and year]
- *29%* were unique based on (gender, dob)
- *69%* were unique (5-digit ZIP, dob)
- *97%* were unique based on (9-digit ZIP, dob)

K-anonymity (Sweeney)

- Focus on *quasi-identifiers*: QI near-unique (or unique) identifiers that are potential for linking.
- Define an equivalence relation on equal values of QI.

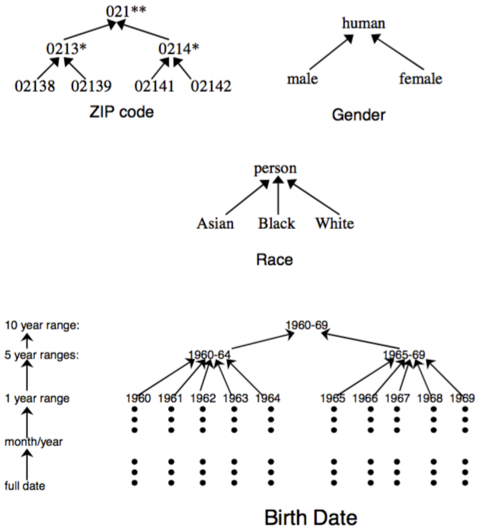
K-anonymity (Sweeney)

- Focus on *quasi-identifiers*: QI near-unique (or unique) identifiers that are potential for linking.
- Define an equivalence relation on equal values of QI.
- Ensure that each equivalence class has size $\geq k > 1$.

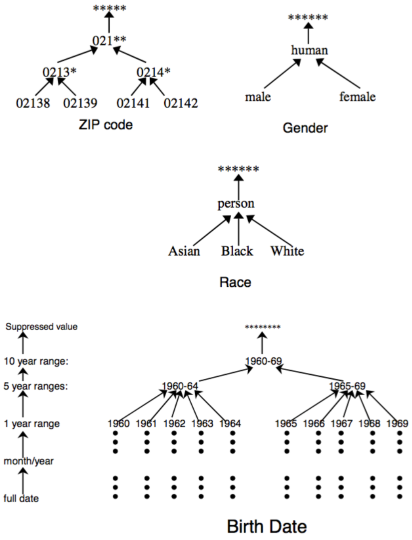
K-anonymity (Sweeney)

- Focus on *quasi-identifiers*: QI near-unique (or unique) identifiers that are potential for linking.
- Define an equivalence relation on equal values of QI.
- Ensure that each equivalence class has size $\geq k > 1$.
- Solution: Release distorted data.
  Distortion through *generalization* and *suppression*.

Generalization Hierarchy



Generalization Hierarchy with Suppression



Example: anonymized table

$$K = 2$$
$$\text{QI} = \{Race,\ Birth,\ Gender,\ Zip\}$$

	Race	Birth	Gender	ZIP	Problem
t1	Black	1965	m	0214*	short breath
t2	Black	1965	m	0214*	chest pain
t3	Black	1965	f	0213*	hypertension
t4	Black	1965	f	0213*	hypertension
t5	Black	1964	f	0213*	obesity
t6	Black	1964	f	0213*	chest pain
t7	White	1964	m	0213*	chest pain
t8	White	1964	m	0213*	obesity
t9	White	1964	m	0213*	short breath
t10	White	1967	m	0213*	chest pain
t11	White	1967	m	0213*	chest pain

Limitations

- Attack: Unsorted matching attack

Limitations

- Attack: Unsorted matching attack
- Attack: Complementary release attack: though order is randomized, link through sensitive attribute.

Limitation: Lack of Diversity

- What if there are K distinct tuples for each QI value, but all of them have the same sensitive values? For example, the K tuples have the same disease.

Race	*DOB*	*Gender*	*ZIP*	*Problem*
White	1965	Male	0214*	Diabetes
White	1965	Male	0214*	Diabetes
White	1965	Male	0214*	Diabetes

Limitation: Lack of Diversity

- What if there are K distinct tuples for each QI value, but all of them have the same sensitive values? For example, the K tuples have the same disease.

Race	*DOB*	*Gender*	*ZIP*	*Problem*
White	1965	Male	0214*	Diabetes
White	1965	Male	0214*	Diabetes
White	1965	Male	0214*	Diabetes

- Attack: another dataset reveals that John, (who visited Mass General), is a white male born in 1965 living in the 02141 zip code

Limitation: Lack of Diversity

- What if there are K distinct tuples for each QI value, but all of them have the same sensitive values? For example, the K tuples have the same disease.

Race	*DOB*	*Gender*	*ZIP*	*Problem*
White	1965	Male	0214*	Diabetes
White	1965	Male	0214*	Diabetes
White	1965	Male	0214*	Diabetes

- Attack: another dataset reveals that John, (who visited Mass General), is a white male born in 1965 living in the 02141 zip code
- Problem: Equivalence class lacks diversity.

[*l*-Diversity: Privacy Beyond *k*-Anonymity. A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. ACM Trans. on Knowledge Discovery from Data, 1 (1), 2007.]

Location based queries

- Find me the nearest restaurant of category C.
  Bob:*I am at* (x_u, y_u)*; I want an item of type* $R.C$ *close to my location.*
- Can such a query reveal sensitive information?

Location based queries

- Find me the nearest restaurant of category C.
  Bob:*I am at* (x_u, y_u)*; I want an item of type* $R.C$ *close to my location.*
- Can such a query reveal sensitive information?
- Attack: Alice can
  1. observe (x_u, y_u) physically;
  2. look up and find it is a house;
  3. apply signal triangulation; ...

Location based queries

- Find me the nearest restaurant of category C.
  Bob:*I am at* (x_u, y_u)*; I want an item of type* $R.C$ *close to my location.*
- Can such a query reveal sensitive information?
- Attack: Alice can
  1. observe (x_u, y_u) physically;
  2. look up and find it is a house;
  3. apply signal triangulation; ...
- Consequences: Consumers may not want
  1. a service that would disclose sensitive personal details;
  2. to be bothered by a deluge of ads.

Location based queries

- Find me the nearest restaurant of category C.
  Bob:*I am at* (x_u, y_u)*; I want an item of type* $R.C$ *close to my location.*
- Can such a query reveal sensitive information?
- Attack: Alice can
  1. observe (x_u, y_u) physically;
  2. look up and find it is a house;
  3. apply signal triangulation; ...
- Consequences: Consumers may not want
  1. a service that would disclose sensitive personal details;
  2. to be bothered by a deluge of ads.
- How to prevent identity inference from location query?

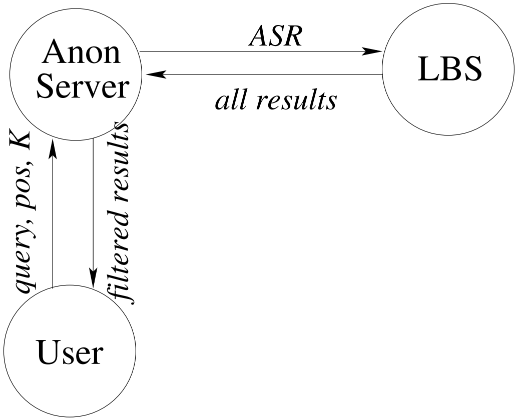
Using K-anonymity

- Location Obfuscation:
- Instead of sending (x_u, y_u), send a region: an *anonymized spatial region (ASR)*.

Using K-anonymity

- Location Obfuscation:
- Instead of sending (x_u, y_u), send a region: an *anonymized spatial region (ASR)*.
- A user enjoys *spatial K-anonymity* in a region if the probability of distinguishing that user from the other users in that region $\leq \frac{1}{K}$.

Setup


Anon Server
ASR
LBS
all results
query, pos, K
filtered results
User

Complications

- How does the Anon Server find the nearby $K - 1$ users?
  - Store user locations in a data structure: quad tree. *(Interval Cloak)*

Complications

- How does the Anon Server find the nearby $K - 1$ users?
  - Store user locations in a data structure: quad tree. *(Interval Cloak)*
- What if $K - 1$ other users cannot be found?

Complications

- How does the Anon Server find the nearby $K-1$ users?
  - Store user locations in a data structure: quad tree. *(Interval Cloak)*
- What if $K-1$ other users cannot be found?
  - Wait until enough users enter! *(Temporal Cloaking)*
    [Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. M. Gruteser and D. Grunwald, MobiSys, 2003.]

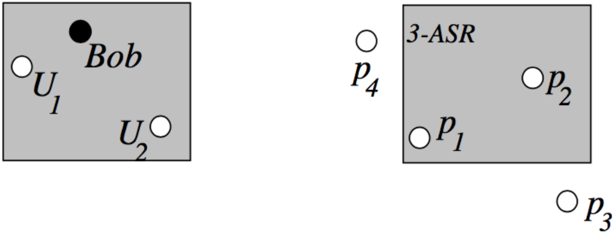
Complications

- How does the Anon Server find the nearby $K-1$ users?
  - Store user locations in a data structure: quad tree. *(Interval Cloak)*
- What if $K-1$ other users cannot be found?
  - Wait until enough users enter! *(Temporal Cloaking)*
    [Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. M. Gruteser and D. Grunwald, MobiSys, 2003.]
  - Create a very large *ASR*. (Bad for LBS.)

Complications

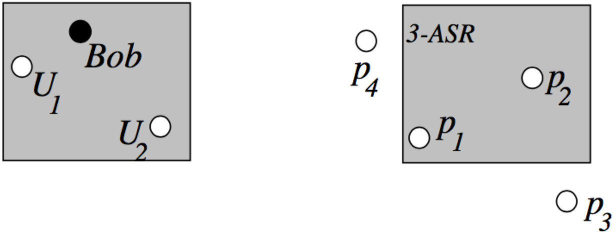
- How does the Anon Server find the nearby $K - 1$ users?
  - Store user locations in a data structure: quad tree. *(Interval Cloak)*
- What if $K - 1$ other users cannot be found?
  - Wait until enough users enter! *(Temporal Cloaking)*
    [Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. M. Gruteser and D. Grunwald, MobiSys, 2003.]
  - Create a very large *ASR*. (Bad for LBS.)
- What does the LBS do? Return the nearest neighbor considering all points in *ASR*.

Nearest neighbor with K-anon



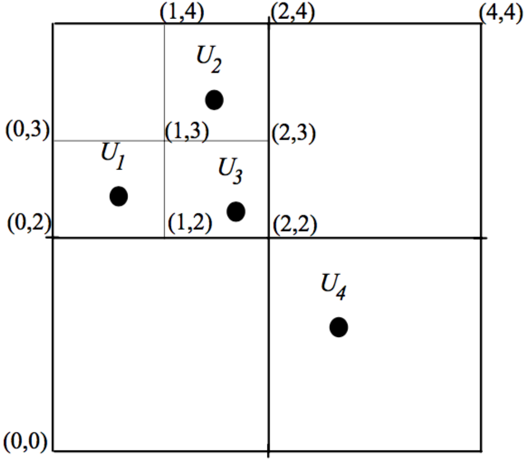
- LBS returns $\{p_1, p_2, p_3, p_4\}$

Nearest neighbor with K-anon



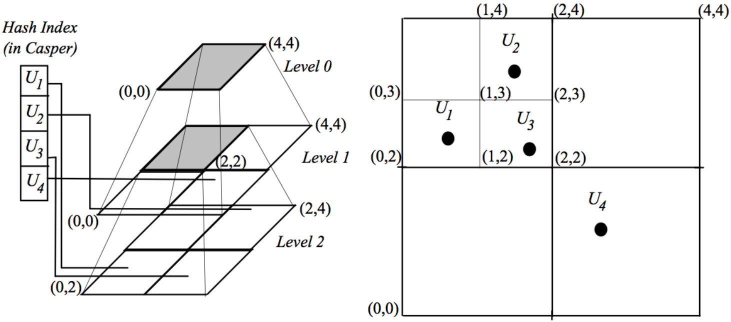
- LBS returns $\{p_1, p_2, p_3, p_4\}$
- Anon Server will filter and return p_2

Example: U_1 wants $K = 2$. ASR?

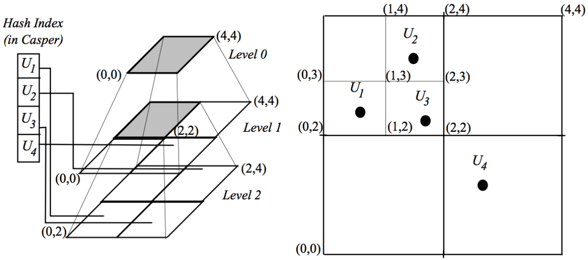


ASR computed is the NW quadrant. (It contains $> K$ users.)

Example: *Casper*



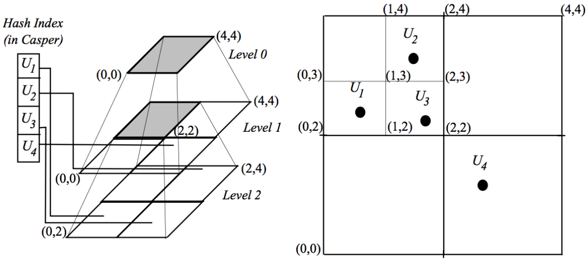
Returns rectangle ((0,2), (2,3))



- Hash index can locate user without search.

[The New Casper: Query Processing for Location Services without Compromising Privacy. M. Mokbel, C. Chow, and W. Aref. VLDB, 2006.]

Casper



- Hash index can locate user without search.
- Searches for neighboring quadrants, i.e., siblings which are neighbors. In this example, we get a smaller ASR.

[The New Casper: Query Processing for Location Services without Compromising Privacy. M. Mokbel, C. Chow, and W. Aref. VLDB, 2006.]

Clique Cloak

- Each query $\Rightarrow$ rectangle with user at centroid. (Dimensions based on heuristics.)

Clique Cloak

- Each query $\Rightarrow$ rectangle with user at centroid. (Dimensions based on heuristics.)
- Create graph $G = (V, E)$:
  $V = \{\text{user-queries}\}$
  $E = \{(u_i, u_j) \mid u_i \text{ falls in } u_j\text{'s rectangle and vice-versa } \}$
  (It is possible for the widths to be different.)

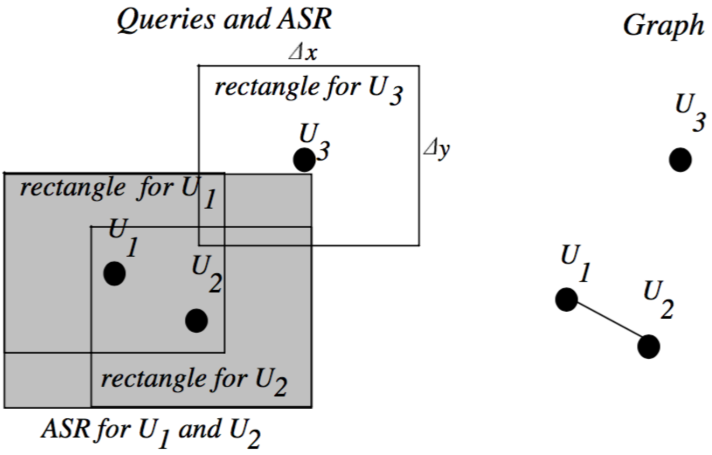
Clique Cloak

- Each query $\Rightarrow$ rectangle with user at centroid.
  (Dimensions based on heuristics.)
- Create graph $G = (V, E)$:
  $V = \{\text{user-queries}\}$
  $E = \{(u_i, u_j) \mid u_i \text{ falls in } u_j\text{'s rectangle and vice-versa } \}$
  (It is possible for the widths to be different.)
- Search G for cliques of size K

Clique Cloak

- Each query $\Rightarrow$ rectangle with user at centroid. (Dimensions based on heuristics.)
- Create graph $G = (V, E)$:
  $V = \{\text{user-queries}\}$
  $E = \{(u_i, u_j) \mid u_i \text{ falls in } u_j\text{'s rectangle and vice-versa } \}$
  (It is possible for the widths to be different.)
- Search G for cliques of size K
- Return minimum bounding rectangle containing all the rectangles corresponding to the clique.

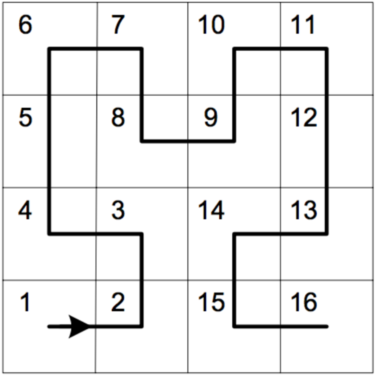
Clique Cloak

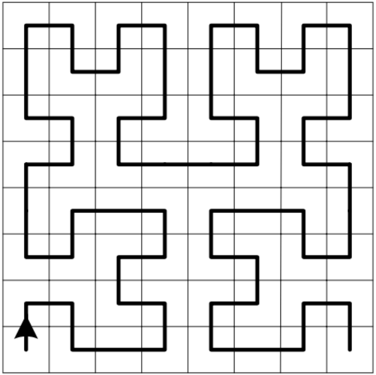


Combined with Temporal Cloaking.

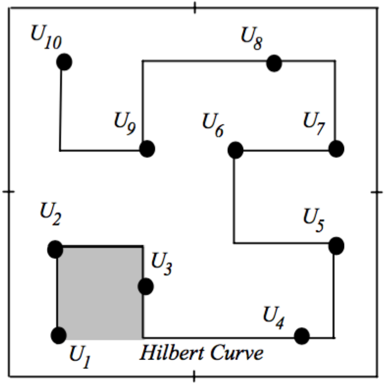
[Location Privacy in Mobile Systems: A Personalized Anonymization Model. Intl. Conf. on Distributed Computing Systems (ICDCS'05) 2005.]

Hilbert Curve





Hilbert Cloak



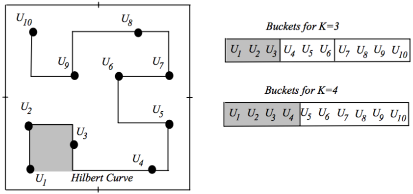
Buckets for K=3

U_1 U_2 U_3	U_4 U_5 U_6	U_7 U_8 U_9 U_{10}

Buckets for K=4

U_1 U_2 U_3 U_4	U_5 U_6 U_7 U_8 U_9 U_{10}

Hilbert Cloak



Kalnis et al.

- Sorts users by their Hilbert cell location;
- Splits sorted list into buckets of K users;
  last bucket may contain up to $K + (K - 1) = 2K - 1$ users.
- Find the bucket corresponding to a user U.
  Return the MBR (min. bounding rect.) for that bucket.

[Preventing Location-Based Identity Inference in Anonymous Spatial Queries.
P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias. IEEE Trans. Knowledge and Data Engineering. 19. 2007.]

Reciprocity

- Suppose user U_1 sends a query with anonymity K and is issued ASR (or anonymizing set AS).
  ASR (AS) satisfies *reciprocity* if it
  1. contains U_1 and $\geq K-1$ other users $U_2 \cdots U_{K-1}$ and
  2. the same ASR is generated for each user $U_2 \cdots U_{K-1}$ for the same K.

Reciprocity

- Suppose user U_1 sends a query with anonymity K and is issued ASR (or anonymizing set AS).
  ASR (AS) satisfies *reciprocity* if it
  1. contains U_1 and $\geq K-1$ other users $U_2 \cdots U_{K-1}$ and
  2. the same ASR is generated for each user $U_2 \cdots U_{K-1}$ for the same K.
- Reciprocity for all AS $\Rightarrow$ Spatial K-anonymity.

Reciprocity

- Suppose user U_1 sends a query with anonymity K and is issued ASR (or anonymizing set AS).
  ASR (AS) satisfies *reciprocity* if it
  1. contains U_1 and $\geq K-1$ other users $U_2 \cdots U_{K-1}$ and
  2. the same ASR is generated for each user $U_2 \cdots U_{K-1}$ for the same K.
- Reciprocity for all AS $\Rightarrow$ Spatial K-anonymity.
- Optimal requirement: ASR should be smallest and reciprocity should be satisfied.
  NP-hard problem.
- Interval Cloak, Casper do not satisfy reciprocity.

Reciprocity

- Suppose user U_1 sends a query with anonymity K and is issued ASR (or anonymizing set AS).
  ASR (AS) satisfies *reciprocity* if it
  1. contains U_1 and $\geq K-1$ other users $U_2 \cdots U_{K-1}$ and
  2. the same ASR is generated for each user $U_2 \cdots U_{K-1}$ for the same K.
- Reciprocity for all AS $\Rightarrow$ Spatial K-anonymity.
- Optimal requirement: ASR should be smallest and reciprocity should be satisfied.
  NP-hard problem.
- Hilbert Cloak achieves reciprocity easily (without optimality).
- Interval Cloak, Casper do not satisfy reciprocity.

Reciprocity

- Suppose user U_1 sends a query with anonymity K and is issued ASR (or anonymizing set AS).
  ASR (AS) satisfies *reciprocity* if it
  1. contains U_1 and $\geq K-1$ other users $U_2 \cdots U_{K-1}$ and
  2. the same ASR is generated for each user $U_2 \cdots U_{K-1}$ for the same K.
- Reciprocity for all AS $\Rightarrow$ Spatial K-anonymity.
- Optimal requirement: ASR should be smallest and reciprocity should be satisfied.
  NP-hard problem.
- Hilbert Cloak achieves reciprocity easily (without optimality).
- Interval Cloak, Casper do not satisfy reciprocity.
- All these algorithms are very efficient for ASR generation.
- At LBS, Interval Cloak is not as efficient.

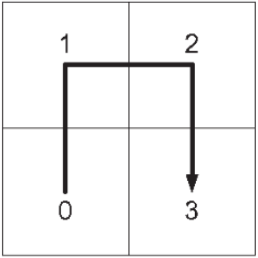
kNN cloak

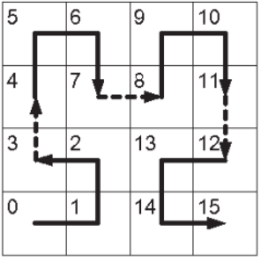
- Preserve K-anonymity and find k nearest neighbors.
  kNN Cloak (Kalnis et al.)

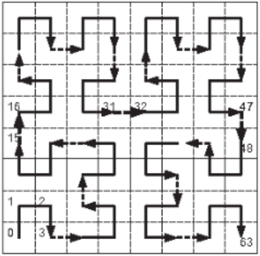
kNN cloak

- Preserve K-anonymity and find k nearest neighbors.
  kNN Cloak (Kalnis et al.)
- Uses R-trees.

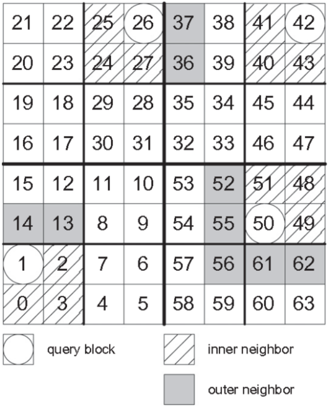
Hilbert Curve: Finding Nearest Neighbors







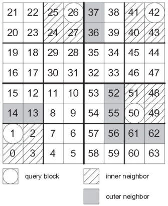
Hilbert Curve: Finding Nearest Neighbors

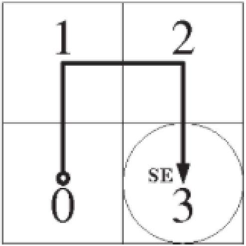


$$\text{Cell } 50 = (3, 0, 2)_4$$

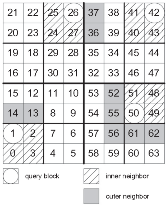
How to find its inner and outer neighbors?

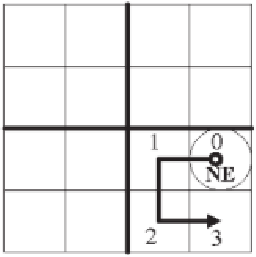
Hilbert Curve: Finding Nearest Neighbors



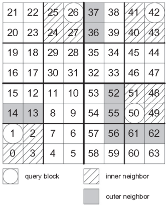


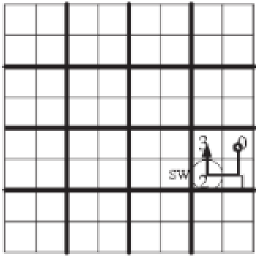
Represent by $(0, 1, 2, \underline{3})$





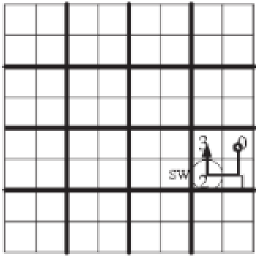
Represent by $(2, 1, \underline{0}, 3)$





Represent by $(\underline{2}, 3, 0, 1)$

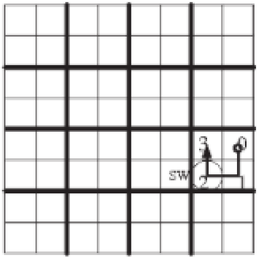
Hilbert Curve: Finding Nearest Neighbors



Represent by $(\underline{2}, 3, 0, 1)$

- At this level, it is the SW cell.
  Neighbor of 50 towards north is NW: +3 -2, i.e., 51
  Neighbor of 50 towards east is SE: +1 -2, i.e., 49

[H. Chen and Y. Chang. All-nearest-neighbors finding based on the Hilbert curve. Expert Systems with Applications. 38(6). 2011.]



Represent by $(\underline{2}, 3, 0, 1)$

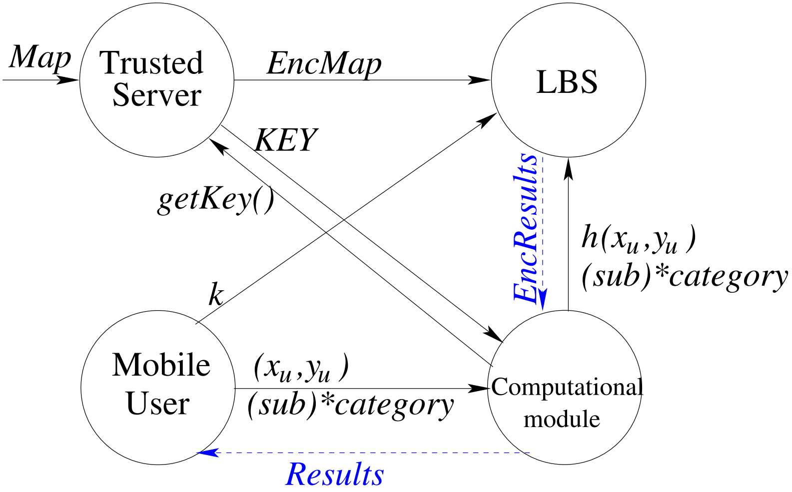
- Neighbor at south is NW cell of $(33)_4$

  ...

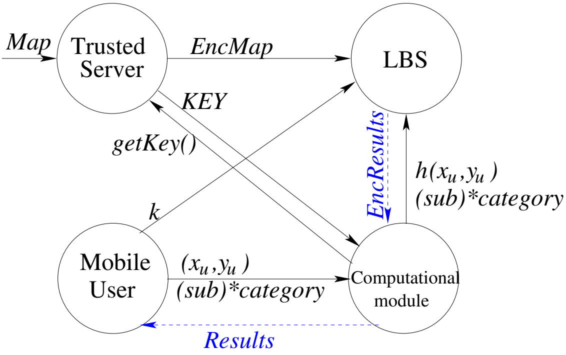
Hiding the Hilbert Curve from the LBS

- Use a single user query without K-anonymity.
- Do not divulge the user location.
- Do not share too much information with any server.
- Hide the Hilbert curve itself from the LBS.

[Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy.
A. Khoshgozaran and C. Shahabi. Trans. Large-Scale Data- and Knowledge-Centered Systems. 2007.]

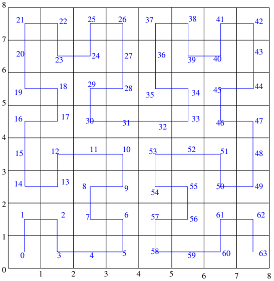

Map
Trusted Server
EncMap
LBS
KEY
getKey()
EncResults
$h(x_u, y_u)$
(sub)*category
k
Mobile User
(x_u, y_u)
(sub)*category
Computational module
Results

Hiding the Hilbert Curve from the LBS



LBS gets a table like

Cell	POI description	Category	Subcategory
43	05A4C3BB02F568489	9A4027D	4715
...	...	...	...
16	47923CC19B6C71AA0	7399BBA	02AA



LBS gets a table like

Cell	POI description	Category	Subcategory
43	05A4C3BB02F568489	9A4027D	4715
...	...	...	...
16	47923CC19B6C71AA0	7399BBA	02AA

Attack

- Guess N

[Can Spatial Transformation-Based Privacy Preservation Compromise Location Privacy? A. Paturi and S. Mazumdar. Intl. Conf. Trust, Privacy and Security in Digital Business (Trustbus) 2018.]

Attack

- Guess N
- Decode a few locations.
  - Look at the category hierarchy;
  - Search for unique instances;
  - Look at clusters and surroundings.

[Can Spatial Transformation-Based Privacy Preservation Compromise Location Privacy? A. Paturi and S. Mazumdar. Intl. Conf. Trust, Privacy and Security in Digital Business (Trustbus) 2018.]

Attack

- Guess N
- Decode a few locations.
  - Look at the category hierarchy;
  - Search for unique instances;
  - Look at clusters and surroundings.
- Find the scaling factor.

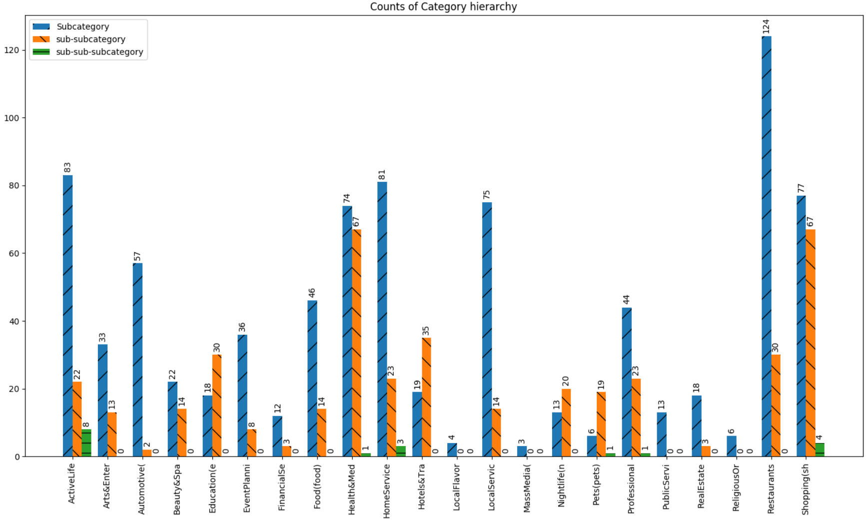
[Can Spatial Transformation-Based Privacy Preservation Compromise Location Privacy? A. Paturi and S. Mazumdar. Intl. Conf. Trust, Privacy and Security in Digital Business (Trustbus) 2018.]

Attack

- Guess N
- Decode a few locations.
    - Look at the category hierarchy;
    - Search for unique instances;
    - Look at clusters and surroundings.
- Find the scaling factor.
- Find which of the rotated/transposed curves it is based on through the relative orientation of 2 decoded locations.

[Can Spatial Transformation-Based Privacy Preservation Compromise Location Privacy? A. Paturi and S. Mazumdar. Intl. Conf. Trust, Privacy and Security in Digital Business (Trustbus) 2018.]

Category Tree: Albuquerque, NM, USA



Single instances

- The *Hotels and Travel* has 19 sub- and 35 subsub-categories. Among those 19 subcategories, there are two with a single instance each: *Airport* and *Ski Resort*.

Single instances

- The *Hotels and Travel* has 19 sub- and 35 subsub-categories. Among those 19 subcategories, there are two with a single instance each: *Airport* and *Ski Resort*.
- How to disambiguate?

Single instances

- The *Hotels and Travel* has 19 sub- and 35 subsub-categories. Among those 19 subcategories, there are two with a single instance each: *Airport* and *Ski Resort*.
- How to disambiguate?
- The airport and the ski resort are on the south and north ends of the city respectively.
  Airport: near a busy freeway; many hotels and restaurants nearby.
  Ski resort: more secluded; surrounded by just a few restaurants.

More generally

- What if users understand and reveal selectively?

Identity, Location

Effectively, adversary gets messages of the form *(Identity, Location)*

Identity	Location	Impact
Hide	Hide	someone was somewhere
Hide	Reveal	someone was here
Reveal	Hide	Alice was somewhere
Reveal	Reveal	Alice was here

[A Classification of Location Privacy Attacks and Approaches. M. Wernke, P. Skvortsov, F. Durr, K. Rothermel. Personal and Ubiquitous Computing. 2013.]

Identity, Location

Identity	Location	Impact
Hide	Hide	somebody was somewhere
Hide	Reveal	someone was here count people in hotel's meeting rooms
Reveal	Hide	Alice was somewhere competitor should not know employee's location
Reveal	Reveal	Alice was here proud of attendance at a tourist location

Adding time

Effectively, adversary gets messages of the form:

(Identity, Location, Timestamp)

Adding Time

- *K-anonymity*: same set of $K - 1$ other users?
  Generalize ASR to spatio-temporal anon. region.

Adding Time

- *K-anonymity*: same set of $K - 1$ other users?
  Generalize ASR to spatio-temporal anon. region.
- *l-diversity*: what if the anonymized spatial region contains exactly one location: a sensitive one?

Adding Time

- *K-anonymity*: same set of $K - 1$ other users?
  Generalize ASR to spatio-temporal anon. region.
- *l-diversity*: what if the anonymized spatial region contains exactly one location: a sensitive one?
- *t-closeness*: what if they are different locations but very similar?

Adding time

Identity	Location	Time	Impact
Hide	Hide	Hide	someone somewhere some time
Hide	Hide	Reveal	someone was somewhere at this time
Hide	Reveal	Hide	someone was here at unknown times time trace without speed info
Hide	Reveal	Reveal	someone was here at these times time trace of people in roads / rooms
Reveal	Hide	Hide	Alice was somewhere at these times protect max speed of known person
Reveal	Hide	Reveal	Alice was somewhere protect sensitive locn. but share presence
Reveal	Reveal	Hide	Alice was here at unknown times Share a visit for potential rescue
Reveal	Reveal	Reveal	Alice was here at known times announce attendance at a tourist location

Strategies

- *Location obfuscation*: send a circular region instead of a position. Map matching reduces its effectiveness: e.g., region contains a lake.

Strategies

- *Location obfuscation*: send a circular region instead of a position. Map matching reduces its effectiveness: e.g., region contains a lake.
- *Spatio-temporal obfuscation*: region with imprecise timestamps allowing K-anonymity to be achieved for each query.

Strategies

- *Location obfuscation*: send a circular region instead of a position. Map matching reduces its effectiveness: e.g., region contains a lake.
- *Spatio-temporal obfuscation*: region with imprecise timestamps allowing K-anonymity to be achieved for each query.
- *Secret Query:* Answer queries without knowing the query: crypto techniques. (Computationally expensive.)

Strategies

- *Location obfuscation*: send a circular region instead of a position. Map matching reduces its effectiveness: e.g., region contains a lake.
- *Spatio-temporal obfuscation*: region with imprecise timestamps allowing K-anonymity to be achieved for each query.
- *Secret Query:* Answer queries without knowing the query: crypto techniques. (Computationally expensive.)
- *Fake queries*;
  Dummies: sequence of fake queries, e.g., from a fake car.

Strategies

- *Location obfuscation*: send a circular region instead of a position. Map matching reduces its effectiveness: e.g., region contains a lake.
- *Spatio-temporal obfuscation*: region with imprecise timestamps allowing K-anonymity to be achieved for each query.
- *Secret Query:* Answer queries without knowing the query: crypto techniques. (Computationally expensive.)
- *Fake queries*;
  Dummies: sequence of fake queries, e.g., from a fake car.
- *Mix Zones:* Send queries only from a zone where the paths of many users intersect, e.g., parking lots in shopping malls; (*Mobimix*)
  Use different pseudonyms each time for the same user.

Dangers

- *Attitudes*

  Denezis et al. asked undergrads (2005) the compensation they needed to share a month's worth of location data.

  Median price was 10 GBP.

  For commercial use of the data, it was double.
  [Danezis, G., S. Lewis, and R. Anderson, How Much is Location Privacy Worth? Fourth Workshop on the Economics of Information Security. 2005: Harvard Univ.]

Dangers

- *Attitudes*
  Denezis et al. asked undergrads (2005) the compensation they needed to share a month's worth of location data.
  Median price was 10 GBP.
  For commercial use of the data, it was double.
  [Danezis, G., S. Lewis, and R. Anderson, How Much is Location Privacy Worth? Fourth Workshop on the Economics of Information Security. 2005: Harvard Univ.]

- Major advances in *facial recognition*.