Privacy Aspects of Location & Mobility Data

Yannis Theodoridis

InfoLab | University of Piraeus | Greece
infolab.cs.unipi.gr

Univ. Cyprus, Nov. 2012

Big Brother is watching you!

G. Orwell, "Nineteen Eighty-Four" (1948)

The Wireless Explosion

Do you use any of these devices?

Have you ever felt that you have been tracked?

From opportunities …

- Our every day actions leave digital traces
  - Shopping transactions with loyalty cards, credit cards etc.
  - Electronic administrative transactions and health records
- Traces are stored because are **worth** being remembered
- Useful applications may be built, and important knowledge may be revealed
  - Mobility data analysis
    - How people move around in the town? Are there typical movement behaviors?
    - How have people movement habits changed in this area during the last decade – year – month – day?

… to threats

- Personal mobility data is extremely sensitive
- Their disclosure may represent a brutal violation of privacy protection rights, i.e., to keep confidential
  - the sensitive places we visit
  - the places we live or work at
  - the people we meet
  - …

The naïve scientist's view

- Knowing the exact identity of individuals is not needed for analytical purposes
  - De-identified mobility data are enough to reconstruct aggregate movement behaviour, pertaining to groups of people.
- Reasoning coherent with European data protection laws: **personal data, once made anonymous, are not subject to privacy law restrictions**
- Is this reasoning correct?

•95/46/EC: Goal is to ensure free flow information. Forbids sharing data with states that don't protect privacy.

•2002/58/EC: protection of analyzing private data and private life in the domain of electronic communication.

Unfortunately not!

- Making data (reasonably) anonymous is not easy.
- Sometimes, it is possible to **reconstruct the exact identities from the de-identified data**.
- Many famous examples of re-identification
  - Governor of Massachusetts' clinical record (Sweeney experiment, 2001)
  - AOL August 2006 crisis: user re-identified fro[illegible]gs
- Two mai[illegible]
  - Many [illegible]
  - Linking [illegible]

Sweeney purchased the voter registration list for Cambridge Massachusetts – 54.805 people.
- 69% of records: unique on zip code and date of birth.
- 87% of records: unique on zip code, date of birth and sex.

Out of them, Sweeney disclosed Governor's clinical record !!

The Sweeney experiment

For example, William Weld was governor of Massachusetts at that time and his medical records were in the GIC data. Governor Weld lived in Cambridge Massachusetts. According to the Cambridge Voter list, six people had his particular birth date; only three of them were men; and, he was the only one in his 5-digit ZIP code.

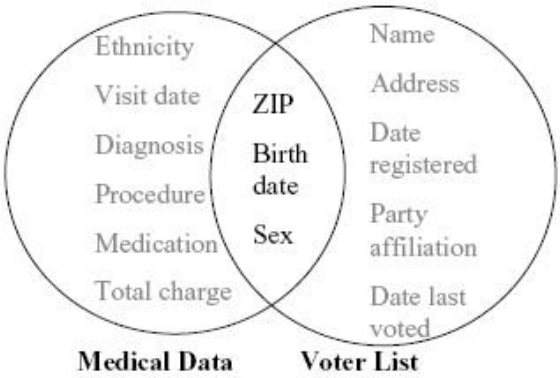


Figure 1 Linking to re-identify data

Source: L. Sweeney. *k*-anonymity: a model for protecting privacy. *Int J on Uncertainty, Fuzziness and Knowledge-based Systems,* 10 (5), 2002; 557-570.

(A parenthesis on K-anonymity in relational databases)

Link Private Information to Person

Date of Birth	Zip Code	Allergy	History of Illness
03-24-79	07030	Penicillin	*Pharyngitis*
08-02-57	**07028**	**No Allergy**	***Stroke***
11-12-39	07030	No Allergy	*Polio*
08-02-57	07029	Sulfur	*Diphtheria*
08-01-40	07030	No Allergy	*Colitis*

Quasi - identifiers

Sensitive **Information**

- **Quasi-identifiers**: a set of attributes that may identify individuals.
- **Sensitive attribute(s)**: information that individuals do not want to be published.

The problem

- Transform ("sanitize") a given dataset so that no one can …
  - … associate a particular record with the corresponding data subject
  - … infer the sensitive information of any data subject
- Transformation must be **minimal** to preserve as much information as possible.
  - Minimize distortion of results.

A solution: K- anonymity

- **K- anonymity** (Sweeney, 2001):
  - **Definition:** A dataset is "K-anonymous" when for any given quasi-identifier, a record is indistinguishable from K-1 other records.
- Practically:
  - Any combination of values of quasi-identifiers should appear at least K times.
  - Under K- anonymity, there will be at least K individuals to whom a given record indistinctly refers.
- The probability that a record belongs to a specific person ≤ 1/K

Techniques for K- anonymization

- **Generalization:** generalize the value to make it less specific.
  - e.g. age “34” becomes “30-40”, zip code 47918 becomes “4791*”
- **Suppression:** simply delete the value
- **Perturbation:** replace the actual value with a random value out of the standard distribution of values for that attribute.
  - The overall distribution of values for that attribute remains the same, but the individual data values are not correct.

Suppression

Original table

Age	Location	Disease
α	β	Flu
α+2	β	Flu
δ	γ+3	Hypertension
δ	γ	Flu
δ	γ-3	Cold

K- anonymized version (K = 2)

Age	Location	Disease
*	β	Flu
*	β	Flu
δ	*	Hypertension
δ	*	Flu
δ	*	Cold

Generalization

Original table

Zip	Gender	Age	Diagnosis
47918	Male	35	Cancer
47906	Male	33	HIV+
47918	Male	36	Flu
47916	Female	39	Obesity
47907	Male	33	Cancer
47906	Female	33	Flu

K- anonymized version (K = 3)

Zip	Gender	Age	Diagnosis
4791*	Person	[35-39]	Cancer
4790*	Person	[30-34]	HIV+
4791*	Person	[35-39]	Flu
4791*	Person	[35-39]	Obesity
4790*	Person	[30-34]	Cancer
4790*	Person	[30-34]	Flu

However ... (Machanavajjhala, 2007)

- K- anonymity is not panacea
  - K- anonymity guarantees that an attacker cannot identify the individual based on the quasi-identifier. But …
- Homogeneity Attacks
  - k-Anonymity is focused on generalizing the quasi-identifiers but does not address the sensitive attributes which can reveal information to an attacker.
- Background Knowledge Attacks
  - Depending on other information available to an attacker, an attacker may have increased probability of being able to determine sensitive information

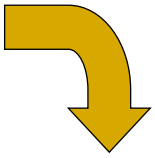
Homogeneity Attacks

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	13053	28	Russian	Heart Disease
2	13068	29	American	Heart Disease
3	13068	21	Japanese	Viral Infection
4	13053	23	American	Viral Infection
5	14853	50	Indian	Cancer
6	14853	55	Russian	Heart Disease
7	14850	47	American	Viral Infection
8	14850	49	American	Viral Infection
9	13053	31	American	Cancer
10	13053	37	Indian	Cancer
11	13068	36	Japanese	Cancer
12	13068	35	American	Cancer

Since Alice is Bob's neighbor, she knows that Bob is a 31-year-old American male who lives in the zip code 13053.

==> Alice knows that Bob's record number is 9,10,11, or 12.

==> Alice knows that Bob has cancer.



Sanitized version

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	130**	< 30	*	Heart Disease
2	130**	< 30	*	Heart Disease
3	130**	< 30	*	Viral Infection
4	130**	< 30	*	Viral Infection
5	1485*	≥ 40	*	Cancer
6	1485*	≥ 40	*	Heart Disease
7	1485*	≥ 40	*	Viral Infection
8	1485*	≥ 40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

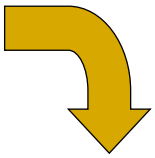
Background Knowledge Attacks

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	13053	28	Russian	Heart Disease
2	13068	29	American	Heart Disease
3	13068	21	Japanese	Viral Infection
4	13053	23	American	Viral Infection
5	14853	50	Indian	Cancer
6	14853	55	Russian	Heart Disease
7	14850	47	American	Viral Infection
8	14850	49	American	Viral Infection
9	13053	31	American	Cancer
10	13053	37	Indian	Cancer
11	13068	36	Japanese	Cancer
12	13068	35	American	Cancer

Alice knows that Umeko is a 21 year-old Japanese female who currently lives in zip code 13068.

==> Alice learns that Umeko's information is contained in record number 1,2,3, or 4.

==> (background knowledge: Japanese have an extremely low incidence of heart disease) Alice concludes with high certainty that Umeko has a viral infection.



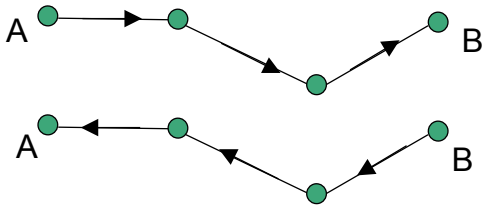
K- anonymous version (K = 4)

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	130**	< 30	*	Heart Disease
2	130**	< 30	*	Heart Disease
3	130**	< 30	*	Viral Infection
4	130**	< 30	*	Viral Infection
5	1485*	≥ 40	*	Cancer
6	1485*	≥ 40	*	Heart Disease
7	1485*	≥ 40	*	Viral Infection
8	1485*	≥ 40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

Back to mobility data …

Linkage in Mobility Data

Id: 34567



[almost every day Mon-Fri between 7:45 – 8:15]

[almost every day Mon-Fri between 17:45 – 18:15]

- By intersecting the phone directories of locations A and B we find that only one individual lives in A and works in B.
- Id:34567 = Prof. Smith
- Then you discover that on Saturday night Id:34567 usually drives to the city red lights district…

Privacy at Different Stages

- Data Collection
  - Privacy Regulations
  - Privacy requirements of different people
  - Sensitivity level depending on context (spatial, temporal, and semantics)
- Data Preprocessing
  - Spatial anonymization is well studied in the context of LBS
  - How about spatio-temporal trajectories?
- Data Analysis

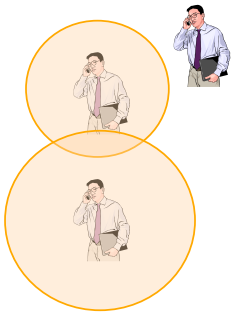
Preserving anonymity

Privacy-preserving mobility data querying

Privacy-preserving mobility data mining

How do people (try to) stay anonymous?

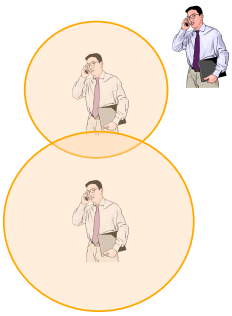
- either by camouflage
  - pretending to be **someone else** or **somewhere else**



- or by hiding in the crowd
  - becoming **indistinguishable** among many others

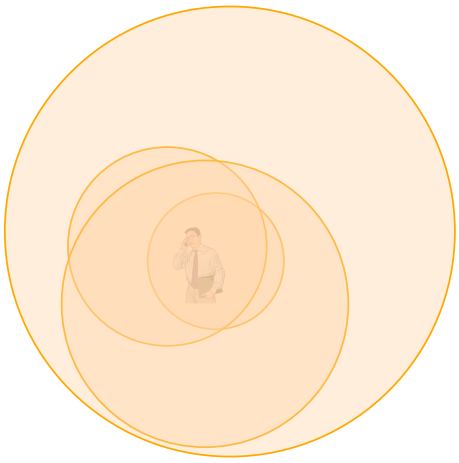
Concepts for Location Privacy

- Location Perturbation – Randomization
  - The user location is represented with a fake value
  - Privacy protection is achieved from the fact that the reported location is false
  - The accuracy and the amount of privacy mainly depends on how far is the reported location from the exact location



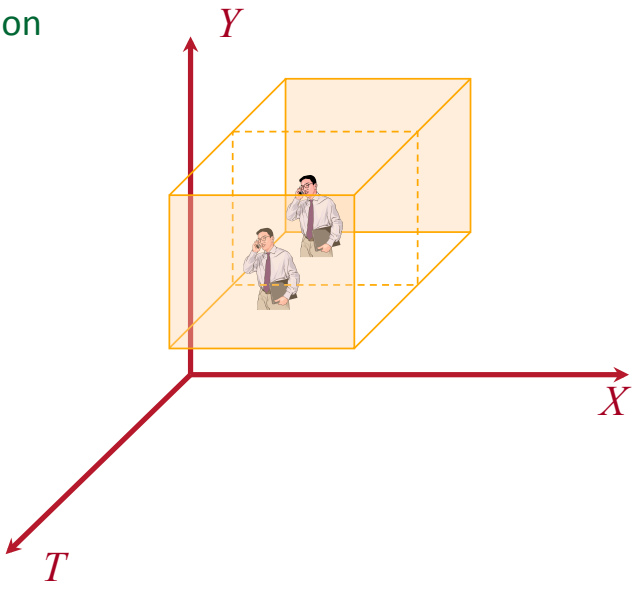
Concepts for Location Privacy

- Spatial Cloaking – Generalization
  - The user exact location is represented as a region that includes the exact user location
  - An adversary does know that the user is located in the region, but has no clue about the exact location
    - The exact location could be anywhere in the region!



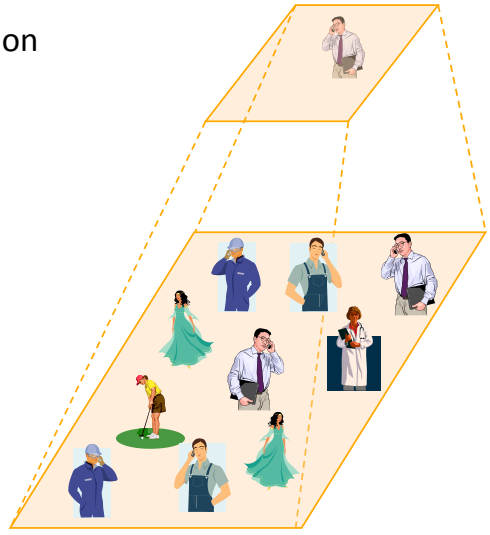
Concepts for Location Privacy

- Spatio-temporal generalization
  - In addition to the spatial dimension, generalize also the temporal dimension



Concepts for Location Privacy

- K- anonymity
  - User's position is generalized to a region containing at least K users
  - The user is indistinguishable among other K-1 users
  - The area largely depends on the surrounding environment.
    - A value of K = 100 may result in a very small area downtown Hong Kong, or a very large area in the desert.



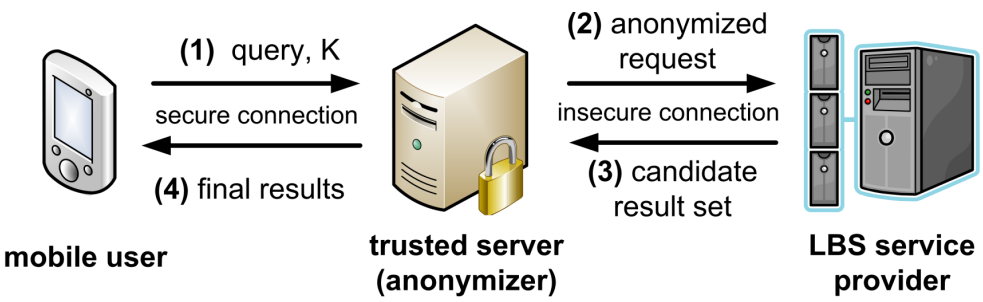
10-anonymity

Privacy in LBS

- Scenario:
  - Bob (an LBS user) asks for a spatial (e.g. NN) query
  - Alice (an attacker) knows the locations of the users. She also "listens" to the responses of the LBS server
    - The LBS server is not trusted!
  - Bob wants to be sure that Alice cannot know his requests

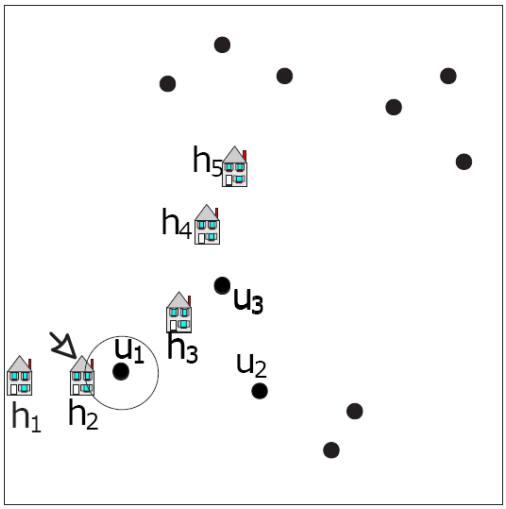
Privacy in LBS

- Scenario:
  - …
  - In order for Bob to be protected by Alice, a trusted **Anonymizer** comes between Bob and LBS server.



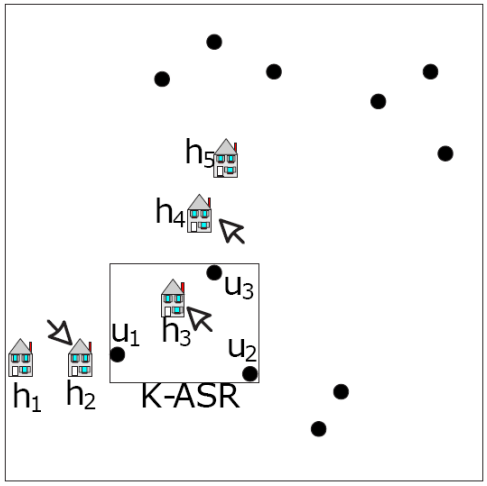
Example of attack in LBS

- Bob asks: "Find the closest hospital to my present location"
- LBS server responds: "h2"
- Alice makes a simple reasoning: it is only u1 that can get h2 as a result!
- Alice implies that Bob asked for the closest hospital
- Solutions to the problem:
  - K- anonymized spatial region (ASR):
    - Casper (Mokbel et al. 2006); Hilbert Cloak (Kalnis et al. 2007)
  - Cryptography
    - PIR (Ghinita et al. 2008)



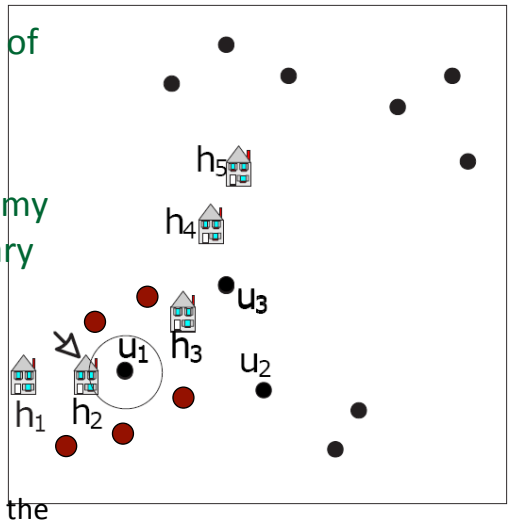
K- anonymized spatial regions

- The location of Bob is generalized to a region that includes at least other K-1 users
  - Bob is indistinguishable among other K-1 users
  - The anonymizer sends the K-ASR to the LBS and asks for the closest hospital
  - The LBS server provides a candidate set of answers
  - The anonymizer sends the correct answer back to Bob
  - Assumption: the value of K is safe! If not?



Introducing dummies … (Kido ICPS '05)

- Introduces several false position data (dummies) along with the true locations of the users to protect the privacy of the requesters of LBSs.
- The challenge is to achieve realistic dummy movements that will confuse an adversary regarding the true locations of the user.
- The location of the first dummies are decided randomly
  - **Moving in a Neighborhood (MN)**: the communication device of the user memorizes the previous position of each dummy. Then the device generates dummies around the memory.

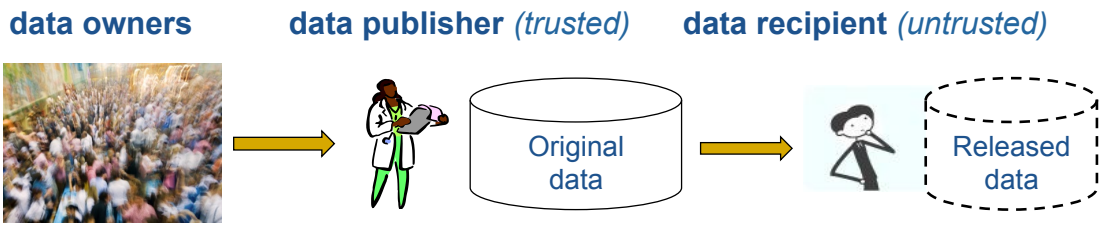


From location data to trajectories

- Pervasiveness of mobility data
- Users' movement traces provide an excellent source of knowledge to aid decision making
- Need to share mobility data for analysis
- Two alternatives for privacy-aware data sharing:
  - Non-interactive scenario (a.k.a. data publishing)
    - **Need an anonymization approach**
  - Interactive scenario (a.k.a. in-house maintenance)
    - **Need a mechanism to regulate the information that is disclosed to (potentially untrusted) end users when querying the database**

Privacy-aware data sharing scenarios

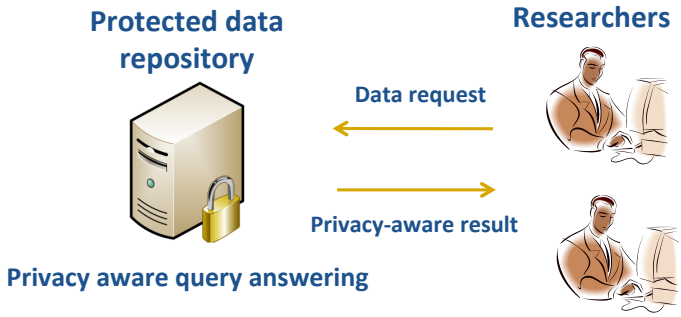
- Non-interactive scenario (a.k.a. data publishing)



- Pros:
  - Constant data availability
  - No infrastructure costs
  - Good for hypothesis generation and testing
  - Seems to model most releases (popular approach)
- Cons:
  - Privacy and utility requirements need to be pre-specified, and the data has to be anonymized accordingly
  - Publisher has no control over the data once they are released
  - No auditing; if a privacy breach occurs there are no means to prevent more (similar) privacy breaches

Privacy-aware data sharing scenarios

- Interactive scenario (akin to statistical databases)



- Pros:
  - Data is kept in-house
  - No need to specify any utility requirements
  - Stronger privacy can be offered
  - Attack identification and recovery from privacy breaches is possible due to auditing
- Cons:
  - Difficulty in answering complex queries
  - Data availability reduces with time since more noise (fakes) is introduced to the DB
  - Infrastructure costs to maintain the DB
  - Bad for hypothesis generation

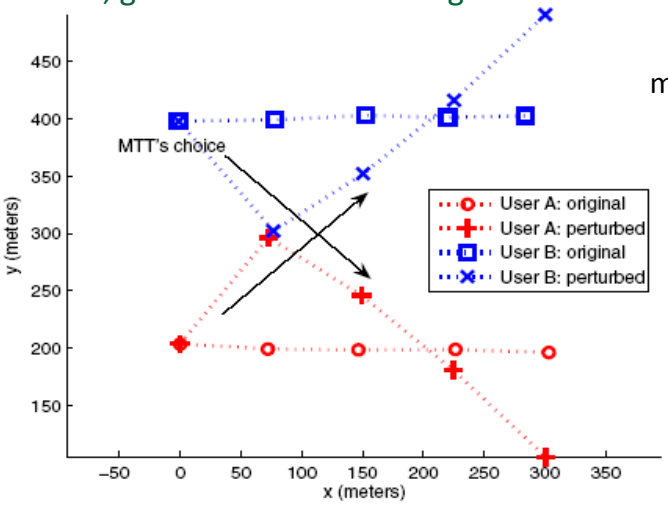
Privacy-preserving MOD mining

- The Perturbation approach:
  - Path confusion: trajectories that are close enough are confused
    - State-of-the-art: Path Crossing [Gruteser & Liu, 2004]
- The K- anonymity approach:
  - Trajectories that are close enough are generalized to be indistinguishable
  - State-of-the-art:
    - Always Walk with Others (AWO) [Nergiz et al. 2008],
    - Never Walk Alone (NWA) [Abul et al. 2008],
    - Wait 4 Me (W4M) [Abul et al. 2010]

Path Confusion

[Gruteser & Liu, 2004]

- Idea of **path crossing**
- When two non-intersecting trajectories are reasonably close to each other, generate a fake crossing of them

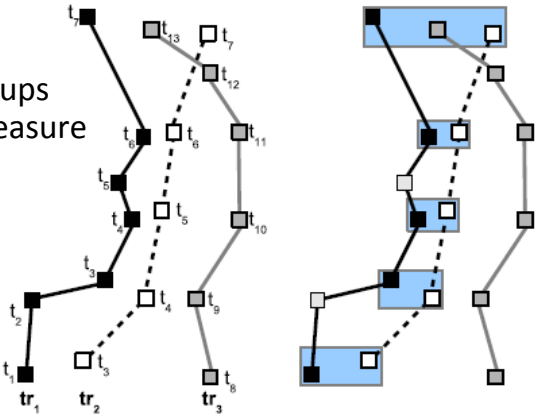


- The radius that indicates the maximum allowable perturbation is the degree of privacy.

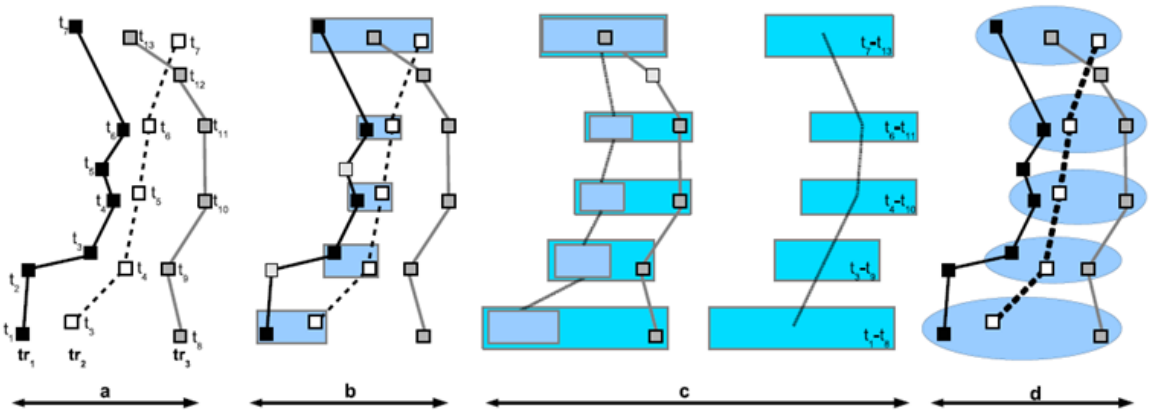
Always Walk with Others (AWO)

[Nergiz et al. 2008]

- Generates a sanitized dataset that consists only of K- anonymous sequences.
- Two steps:
  - Step 1: cluster trajectories into groups of size > K, based on a similarity measure (cost optimal anonymization)
  - Step 2: compute a matching point between the points of the pairs of the trajectories that have been clustered together. Replace the matched points by their MBR. Suppress the unmatched points.



AWO example



a. three trajectories, tr_1, tr_2, tr_3

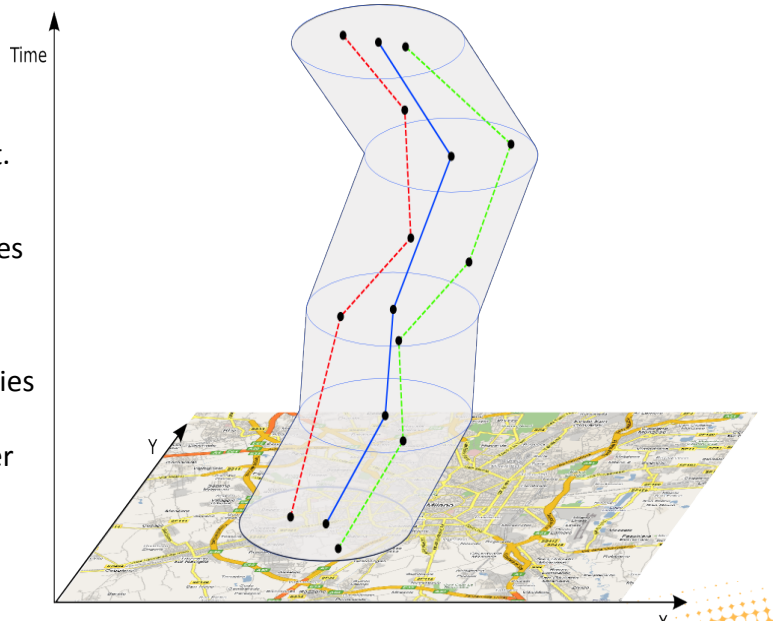
b. and c. anonymization of the three trajectories (in two steps)

d. point matching (5 point links) used in the anonymization of the three trajectories

Never Walk Alone (NWA)

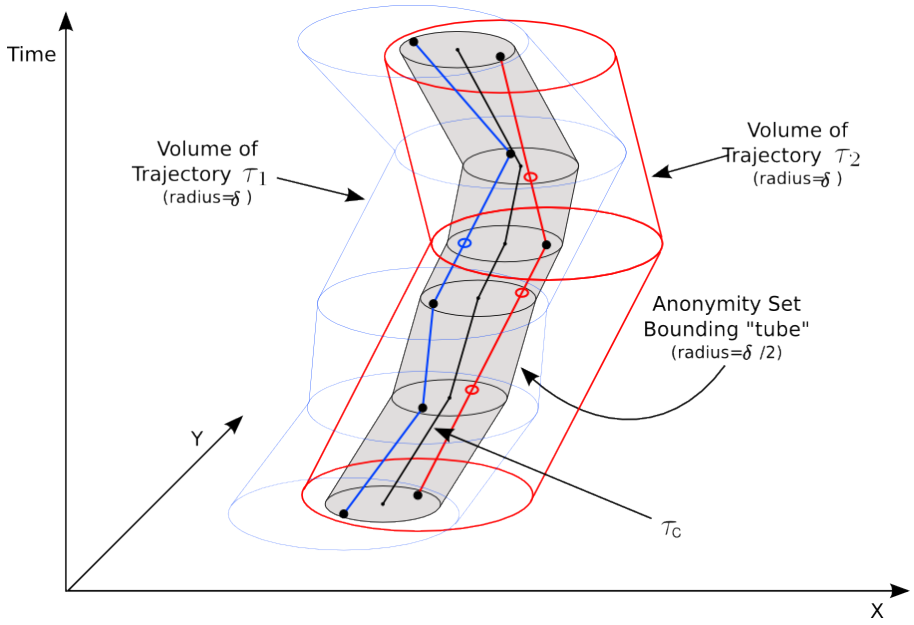
[Abul et al. 2008]

- Three steps:
  - Step 1: Split MOD into equivalence classes w.r.t. trajectories' lifespan
  - Step 2: cluster trajectories into groups of K similar ones (removing outliers)
  - Step 3: perturb trajectories in a cluster so that each one is close to each other up to the uncertainty threshold – (K,δ)- anonymity set



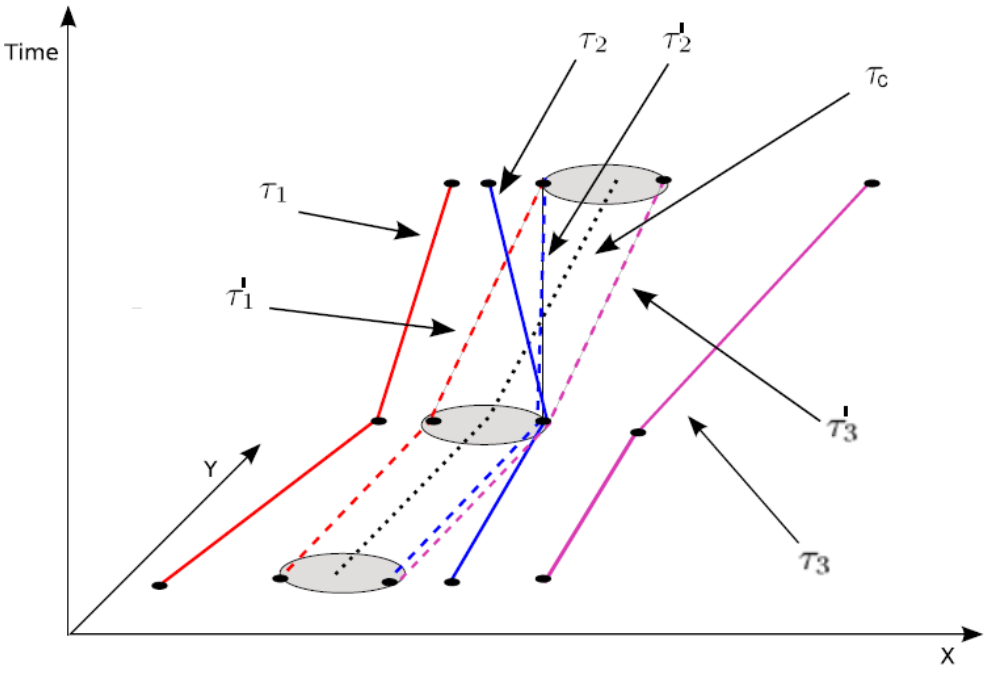
(K,δ)– anonymity set

- K = minimum number of trajectories in the set
- δ = uncertainty threshold (e.g., measurement error of GPS device)



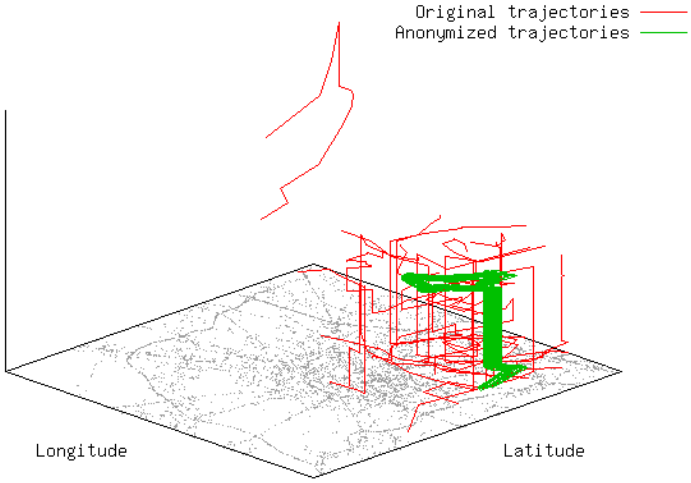
Space translation

- For each cluster, find cluster center, and apply space translation

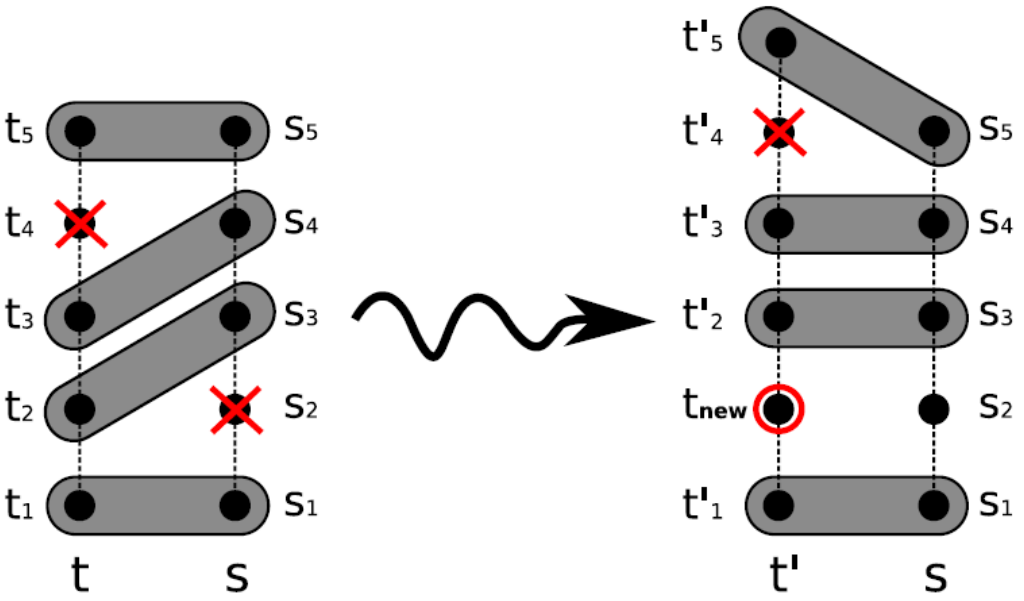


Wait for Me (W4M)

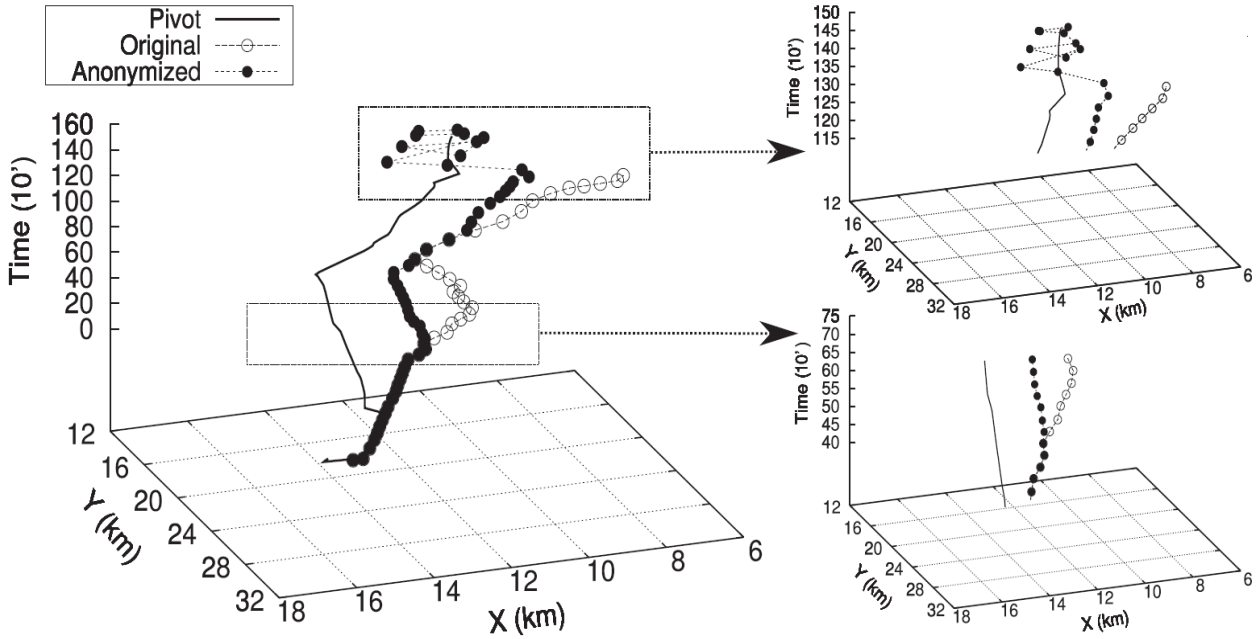
- Variant of NWA with the following improvements
  - Uses a time-tolerant distance measure (i.e. EDR)
  - No need to group into equivalence classes w.r.t. lifespan
  - Spatio-temporal editing instead of space translation



From EDR matching to ST-editing

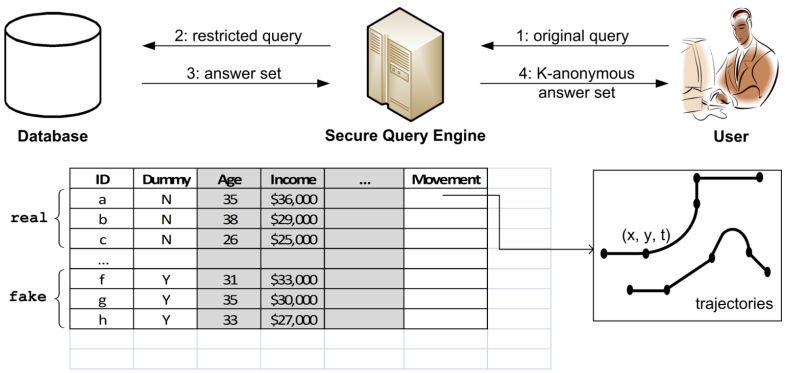


Effect of ST-editing



Privacy-aware MOD querying

- Envisioned privacy-aware query engine [Gkoulalas-Divanis et al. 2008]:

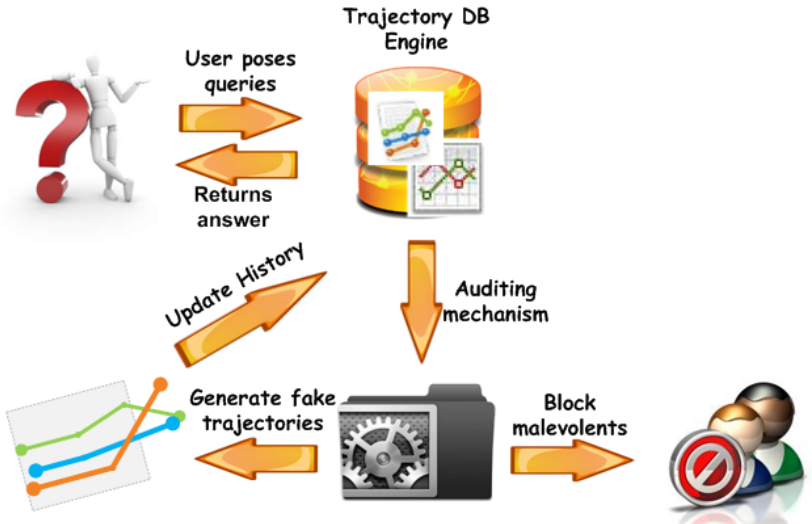


- Allows subscribed end-users to gain restricted access to the database to accomplish various analysis tasks
- Each transaction refers to a person and may contain both trajectory (movement) and non-trajectory (relational) data

Hermes++: A privacy-aware query engine

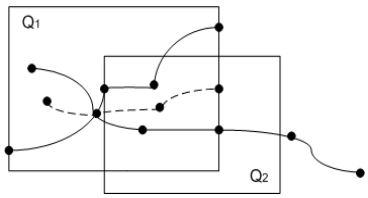
- Audits queries for trajectory data to block potential attacks to user privacy:
  - User identification attack
  - Sensitive location tracking attack
  - Sequential tracking attack
- Supports range, distance, and k-nearest neighbors spatial and spatiotemporal queries
- Preserves user anonymity in answers to queries by:
  - replacing the real trajectories with a set of carefully crafted, realistic fake trajectories, and
  - ensuring that no user-specific sensitive locations are reported as part of the returned trajectories

The "Big Picture" of the proposed engine...

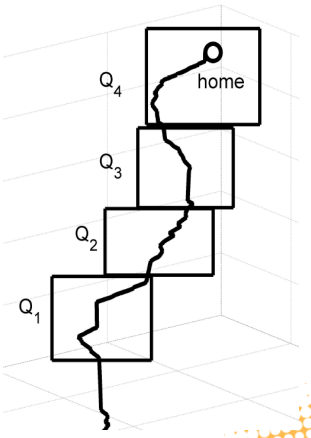


Privacy attacks handled by HERMES++

- User Identification Attack
  - Try to: Expose the user identity by ad hoc queries
  - How? Each query involves the attributes—values pairs of previous queries along with new pairs that make it more specific

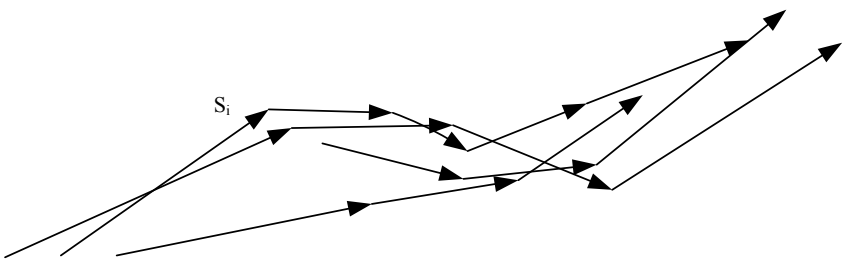


- Sensitive location tracking attack
  - Try to: Identify the starting and/or the ending location of a user trajectory or a user-specific sensitive location
  - How? Map matches the location with a known origin or destination, or with a sensitive place
- Sequential Tracking Attack
  - Try to: Follow a specific user in the system through his/her trajectories to learn the locations that he/she visited
  - How? Focused queries on adjacent regions in space & time



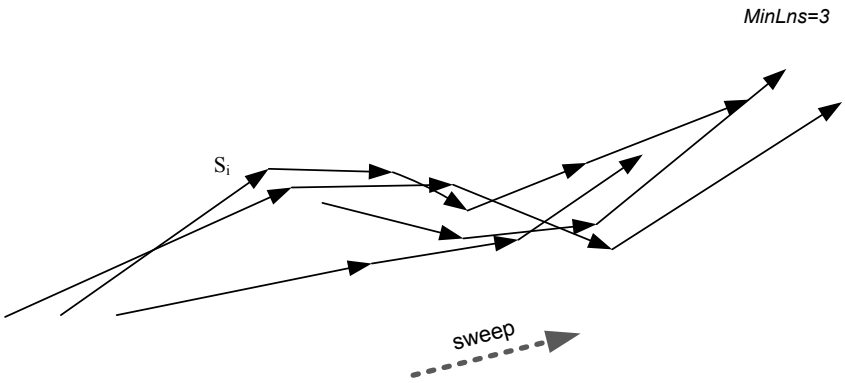
Fake Trajectory Generation

- The idea: **produce fake trajectories for different query types that follow the trend of the result set of real trajectories**
- It is used by our auditing mechanism to handle attacks
- Extends Representative Trajectory Algorithm (RTG) [Lee et al. 2007] by integrating the time dimension



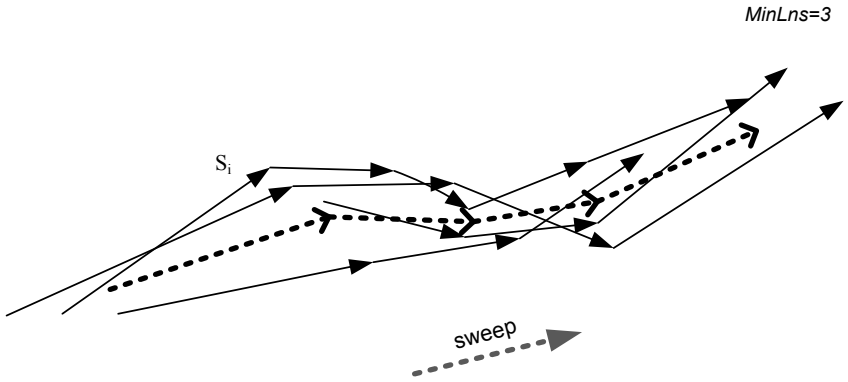
Fake Trajectory Generation

- The idea: **produce fake trajectories for different query types that follow the trend of the result set of real trajectories**
- It is used by our auditing mechanism to handle attacks
- Extends Representative Trajectory Algorithm (RTG) [Lee et al. 2007] by integrating the time dimension



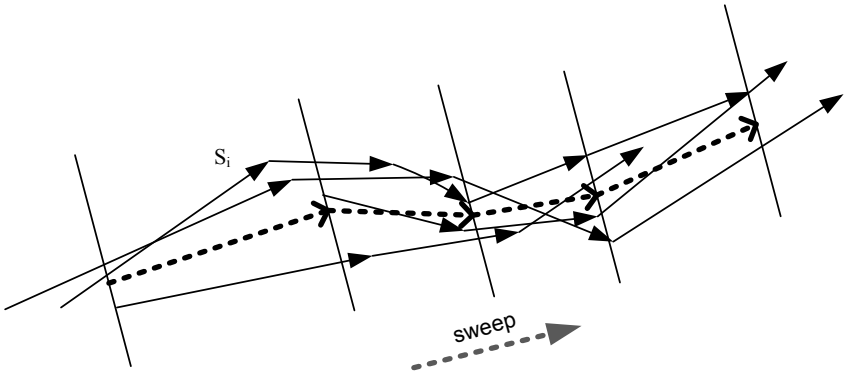
Fake Trajectory Generation

- The idea: **produce fake trajectories for different query types that follow the trend of the result set of real trajectories**
- It is used by our auditing mechanism to handle attacks
- Extends Representative Trajectory Algorithm (RTG) [Lee et al. 2007] by integrating the time dimension



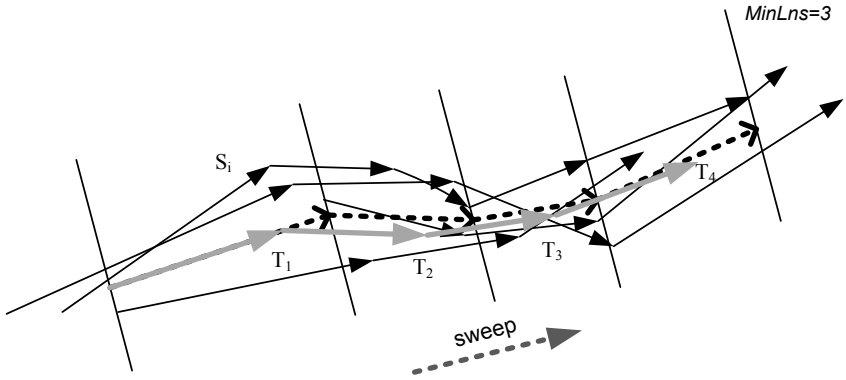
Fake Trajectory Generation

- The idea: **produce fake trajectories for different query types that follow the trend of the result set of real trajectories**
- It is used by our auditing mechanism to handle attacks
- Extends Representative Trajectory Algorithm (RTG) [Lee et al. 2007] by integrating the time dimension



Fake Trajectory Generation

- The idea: **produce fake trajectories for different query types that follow the trend of the result set of real trajectories**
- It is used by our auditing mechanism to handle attacks
- Extends Representative Trajectory Algorithm (RTG) [Lee et al. 2007] by integrating the time dimension



Outline of the *FAKE-GEN* algorithm

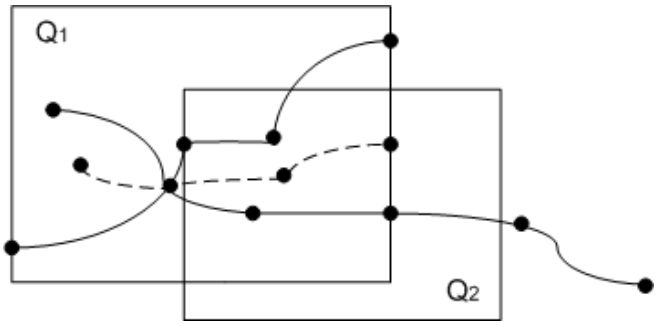
- Operates on 3D directed segments that follow (more or less) the same direction
- Uses *RTG* to create a Representative Trajectory
- Calculates "rational" initial/final timestamps through line simplification
- Uses the statistics of the input 3D directed segments
  - Adjusts the length that a segment can have in order to be more realistic
  - Sets the timestamps of the initial and final point of each segment (regular sampling)
  - If the speed of a segment is not within "rational" limits, it changes the coordinates of the ending position
- Map-matches the generated fake trajectory on a road network

High level design decisions

- Provides answers only if at least *L* real user trajectories exist in the result
- Lower bound *L:*
  - Prevents answering queries whose original result set is very small
  - Avoids failure of capturing the trend of the real trajectories
- Generates *N* fake trajectories,
  - *N* is an owner-specified threshold
- Storage of fakes and past queries of users in the MOD
- Uses spatiotemporal auditing to block privacy attacks
  - Template auditing algorithm for all types of queries (i.e. range, k-NN)

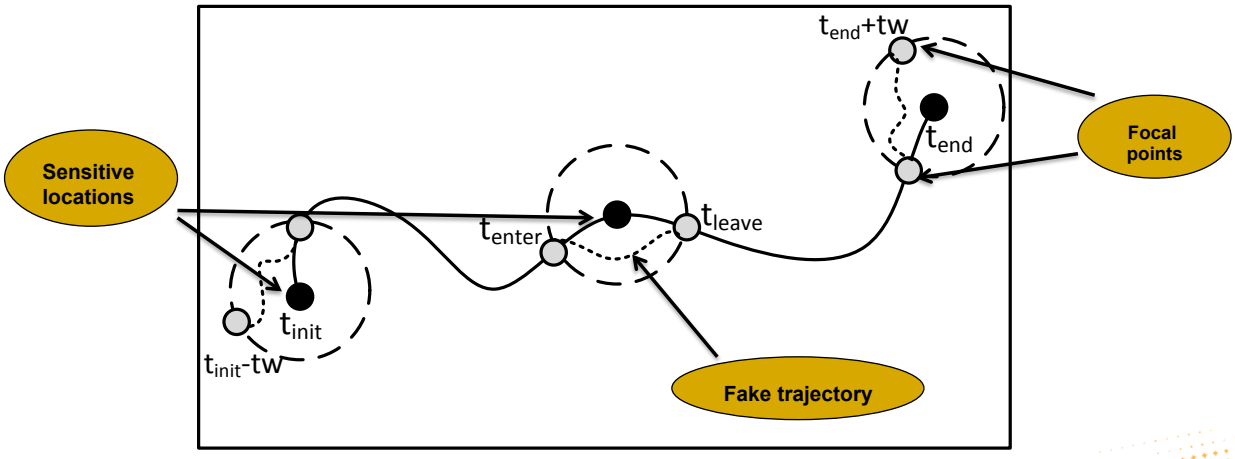
Query Auditing Algorithm-*TrajAuditor* (1/5)

- User identification attack
  - Denies servicing overlapping queries, submitted by the same end-user



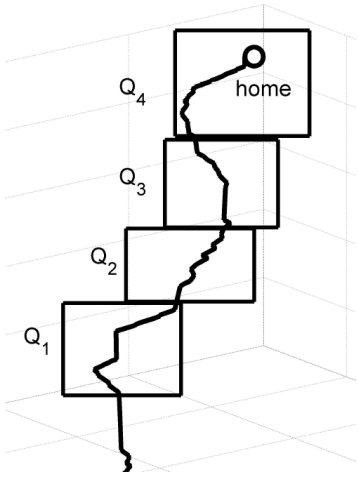
TrajAuditor (2/5)

- Sensitive location tracking attack
  - Prohibits linking sensitive locations visited by a user
    - *'Hide Sensitive Location Algorithm' based on* GSTD* synthesizer



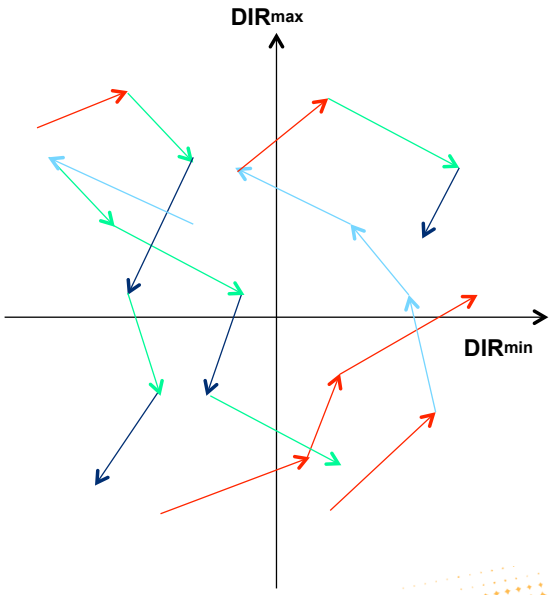
TrajAuditor (3/5)

- Sequential tracking attack
  - The user is tracked down by a set of focused queries on nearby regions (w.r.t. a spatial, temporal threshold)



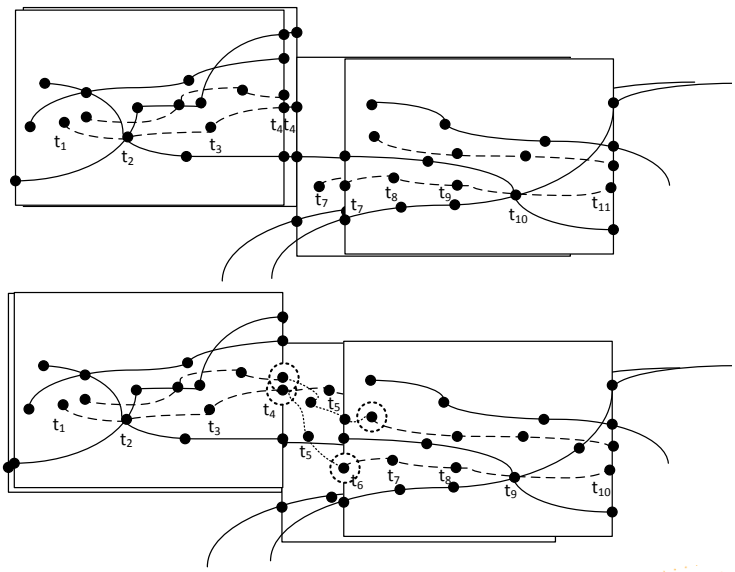
TrajAuditor (4/5)

- Generation of fake trajectories
  - Plane splitting in equivalence classes w.r.t. directionality of input segments
  - Selects segments that follow more or less the same direction (i.e. from one equivalence class)
  - *Fake-Gen* algorithm is applied on such a class
  - The procedure is repeated for the next range of directions until the requested numbers of fakes reaches *N*



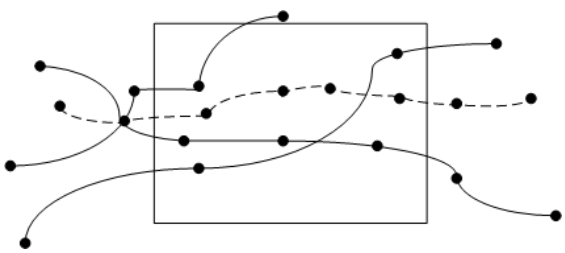
TrajAuditor (5/5)

- Fakes' consolidation for nearby queries (w.r.t a spatial, temporal threshold)
- Performs a one-by-one matching
  - Case I: a space time translation is performed to connect the two fake trajectories
  - Case II: generates a connection-trajectory (dotted lines) between them by using the GSTD*



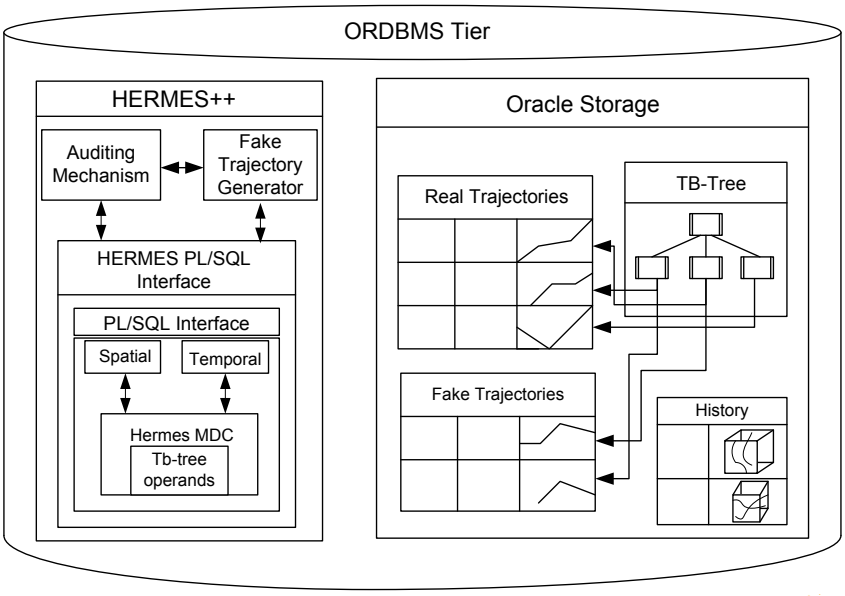
Alternative approach - *TrajFaker*

- TrajFaker vs. TrajAuditor
  - Finds the trajectories that are contained in the query window
  - But … retrieves (and applies *Fake-Gen* to) the whole trajectories
  - Each generated fake trajectory (dotted line) is examined only if it crosses the spatiotemporal window of the query
  - Fake trajectories are stored in order to participate in the generation of other fakes



HERMES ++: System Architecture

- System extension of the HERMES MOD engine
- Built at the ORDBMS level (queries are posed through PL/SQL interface)
- Real, fake trajectories and users' history data are stored in the MOD



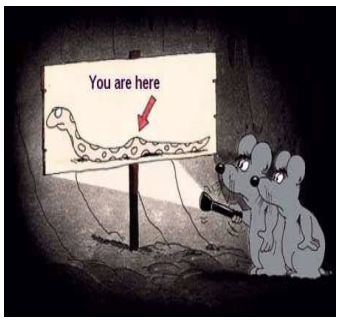
Wrapping up on Hermes++

- Hermes++ follows a "conservative" behavior
  - Data stay in-house
- Supports a variety of popular spatial and spatiotemporal queries
  - The resulted API is simple SQL
- Uses auditing and fake trajectory generation techniques to identify and block, potential attacks to user privacy

Summary

Summary on Privacy Aspects

- Today, tracking of users' locations is an every time / everywhere process
  - Therefore, privacy-preservation is a must!
- What has been done / is to be done:
  - Privacy-aware LBS
    - Several nice solutions are out there.
  - Privacy-aware MOD management
    - Not so many solutions …
  - Privacy-aware KDD process
    - Apart from new techniques …
    - … a theoretical framework is still missing!



Open issues

- K-anonymity
  - What should be the value of this famous "**K**" ?
  - Define an acceptable formal measure of anonymity protection:
    - Probability of re-identification (in a given context)
- Sampling: a necessity and an opportunity!
  - Necessary for performance/feasibiliy of data mining from massive mobility datasets
  - Good for anonymity (re-identification probability decreases)

Questions

Reading list

Privacy-preserving MOD / LBS

- Gedik, B. and Liu, L. (2005) Location Privacy in Mobile Systems: A Personalized Anonymization Model. Proceedings of ICDCS.
- Ghinita, G. et al. (2008) Private Queries in Location Based Services: Anonymizers are not Necessary, Proceedings of ACM SIGMOD.
- Gkoulalas-Divanis, A., and Verykios, V.S. (2008) A Privacy-Aware Trajectory Tracking Query Engine. SIGKDD Explorations, 10(1): 40-49.
- Gkoulalas-Divanis, A. et al. (2009). PLOT: Privacy in Location-Based Services: an Open-Ended Toolbox. Proceedings of MDM.
- Gruteser, M. and Liu. X. (2004) Protecting Privacy in Continuous Location-tracking Applications. IEEE Security and Privacy, 2(2): 28-34.
- Kalnis, P. et al. (2007) Preventing Location-Based Identity Inference in Anonymous Spatial Queries. IEEE Trans. Knowledge & Data Engineering, 19: 1719-1733.
- Kido, H. et al. (2005) An anonymous communication technique using dummies for location-based services. Proceedings of ICPS.
- Mokbel, M.F. et al. (2007) The New Casper: A Privacy-Aware Location-Based Database Server. Proceedings of ICDE.
- Pelekis et al. (2011). Privacy-Aware Querying over Sensitive Trajectory Data, Proceedings of CIKM.

Privacy-preserving MOD mining

- Abul, O. et al. (2007a) Hiding Sensitive Trajectory Patterns. Proceedings of ICDM Workshops.
- Abul, O. et al. (2007b) Hiding Sequences. Proceedings of ICDM Workshops.
- Abul, O. et al. (2008) Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases. Proceedings of ICDE.
- Terrovitis, M. and Mamoulis, N. (2008) Privacy Preservation in the Publication of Trajectories. Proceedings of MDM.
- Zacharouli, P. et al. (2007) A K-Anonymity Model for Spatiotemporal Data. Proceedings of STDM.