AALBORG UNIVERSITY

Location Privacy in LBS (Part I)

Simonas Šaltenis

daisy.cs.aau.dk

Outline

- Location Privacy – An Overview*
  - Assumptions, requirements, and challenges
  - Location privacy problems (attacks on privacy)
  - High-level overview of the proposed solutions

- *G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan, "Private Queries in Location-Based Services: Anonymizers are Not Necessary," ACM SIGMOD 2008***

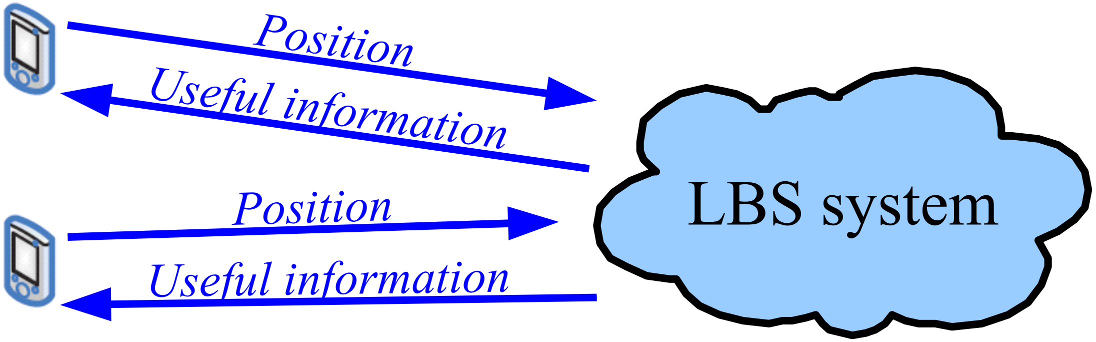
* Based on *M.Decker "Location Privacy – An Overview," 7th IEEE Intl. Conf. On Mobile Business, 2008.*

Acknowledging material from Ling Liu's (VLDB 2007 tutorial) and M.F. Mokbel (VLDB 2006 paper) slides.

** Acknowledging material from P.Kalnis slides

Location Based Services

- Location Based Services (LBS)
  - Internet services (usually mobile) that use geo-location(s) of the user(s) to provide services
    - Example: “Nearest restaurant” service
  - Geo-Location:
    - Current location (+ velocity vector)
    - Past locations
    - Locations of other users
      - “track-my-kid” and “friend-finder” services



LBS: Example Queries

- Location-based emergency services & Traffic monitoring:
  - *Range query*: *How many cars on the highway E-45 north in Aalborg?*
  - *Nearest-neighbor query*: *Give me the location of 5 nearest Toyota maintenance stations?*
- Location-based advertisement/entertainment:
  - *Range query*: *Send E-coupons to all customers within five miles of my store*
  - *Nearest-neighbor query*: *Where is the nearest movie theater to my current location?*
- Other “Points of Interest” (POI) location services:
  - *Range query*: *Where are the gas stations within five miles of my location?*
  - *Nearest-neighbor query*: *Where is the nearest grocery store?*

Privacy

- Location privacy – the claim/right of individuals, groups and institutions to determine for themselves, *when*, *how* and to *what extent* location information about them is communicated to others
  - Part of a more general concept of data privacy

- Location privacy is in conflict with context awareness – using all the available information about the user's context (including its location) to provide a relevant, unobtrusive service.

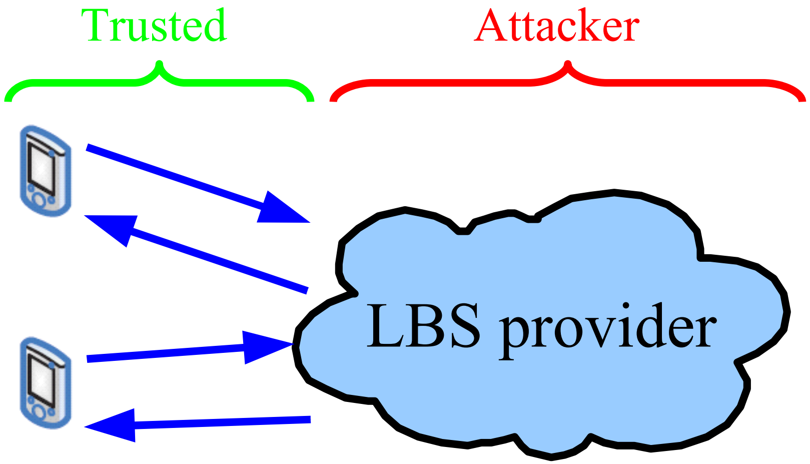
- Important – *assumptions* (not always clearly stated):
  - What exactly is the object of privacy?
  - Who is the attacker and what knowledge is available to the attacker?

Key Assumptions

- Different geo-positioning technologies:
  - Client-based positioning (GPS, Galileo)
  - Network-based positioning (cellular networks, in-door positioning)
- Assumption: the source of geo-locations is trusted.

- An *attacker* is the LBS provider (or someone who compromised the provider's systems)
  - An *attack* is successful, when LBSP gains more knowledge about a user's location(s) than the user intended to let the LBSP know.

  - Client hardware and communication links are considered trustworthy and not compromised

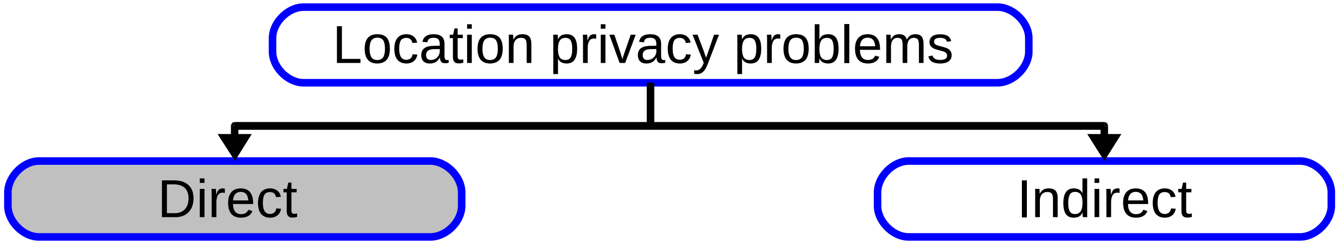


Challenge – Query Processing

- *Why not just encrypt information*?
  - The LBS server needs to process queries!

- Three cases [Mokbel et al., VLDB 2006]:
  - Private queries over public data ← *Most research*
    - *What is my nearest gas station*?
  - Public queries over private data
    - *How many cars in on the E45 north in Aalborg*?
  - Private queries over private data
    - *Where is my nearest friend*?

Location Privacy Problems I



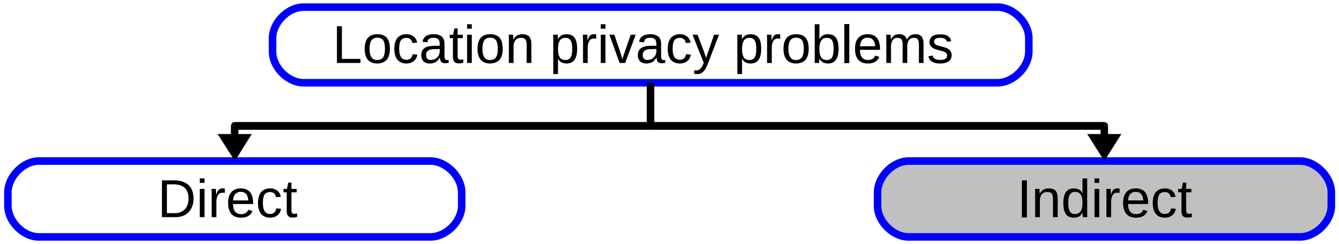
Location privacy problems

Direct

Indirect

- *Direct* location privacy problems:
  - Knowing *where.* Knowing that *Alice* has visited location *L* may reveal:
    - Political, religious, etc. views (party headquarters, church)
    - Personal interests (shops, clubs...)
    - Employer
    - Circle of friends (friend's house)
    - Health problems (hospital)
  - Knowing *when.* Knowing that *Alice* has visited location *L* at time *T.*
  - Knowing *how many times.* Knowing the history (L_1, T_1), (L_2, T_2), ... , (L_n, T_n)

Location Privacy Problems II



Location privacy problems

Direct

Indirect

- Some LBS may not need to know the user's true identity. Thus, pseudonymization can be applied
  - A *mediator* replaces the user's identity by a pseudonym in each request to the LBS provider.

- *Indirect* privacy problems involve attacks on pseudonyms
  - Location information + other external information = revealed identity of the user

Attacks on Pseudonyms

- Known-place attack.
  - External information = knowledge about places where certain users typically stay (e.g., work, home address from public telephone books)

- Commuter attack.
  - Like the *known-place attack*, but based on a recorded spatio-temporal track of requests.

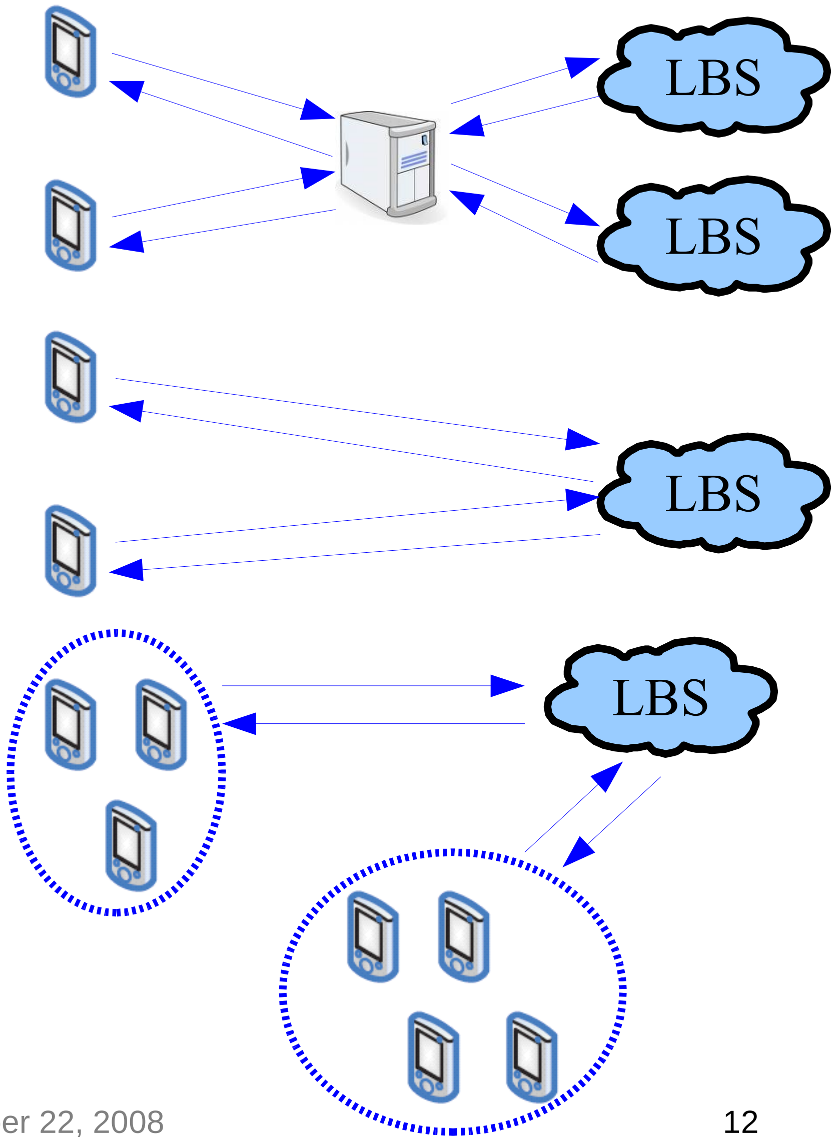
- Observation attack.
  - External information from observation cameras, car number plate recognition systems enables to correlate (through a shared location) known identity with a pseudonym.

Outline

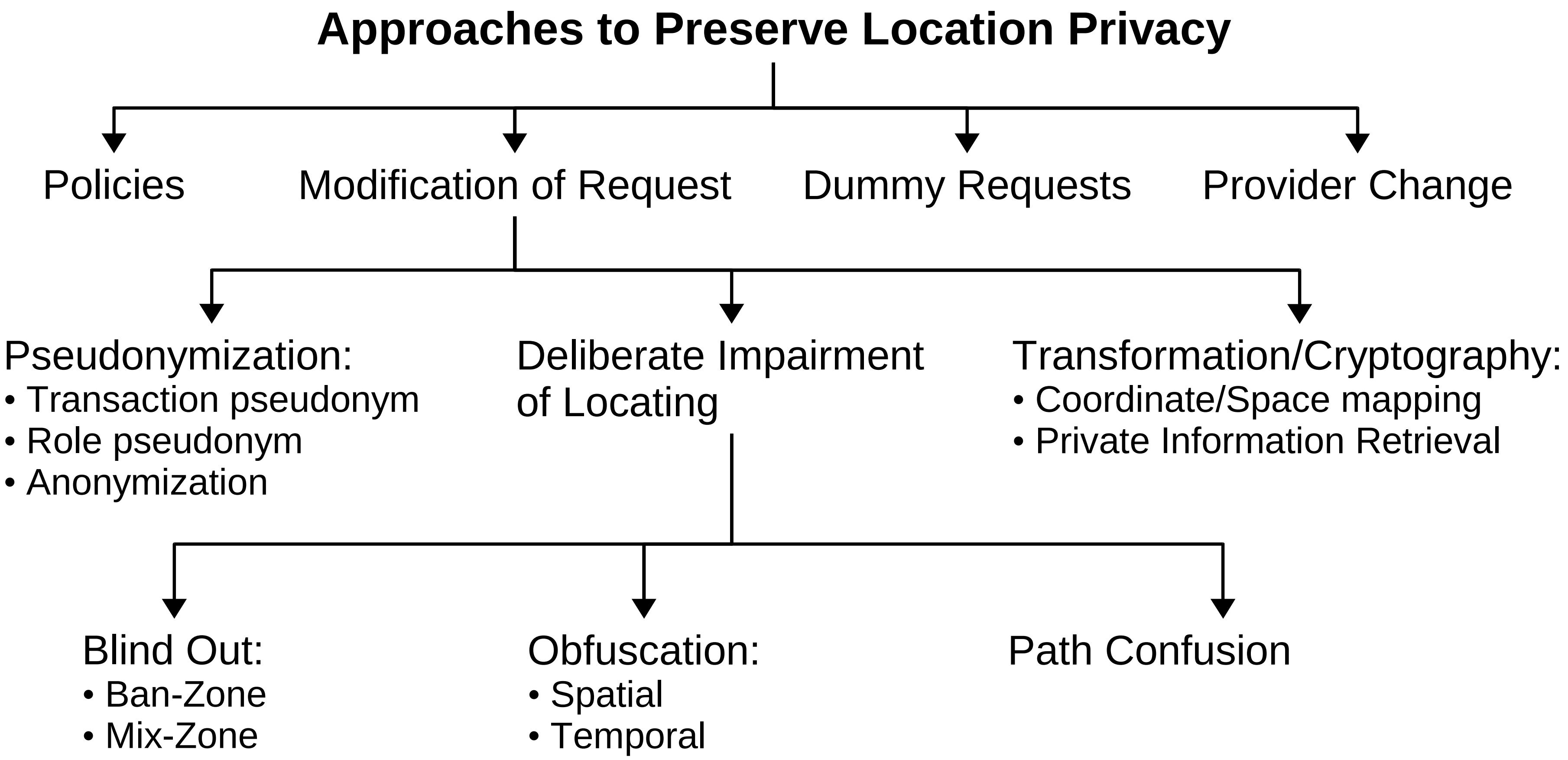
- Location Privacy – An Overview
  - Assumptions, requirements, and challenges
  - Location privacy problems (attacks on privacy)
  - High-level overview of the proposed solutions

Architectures

- Possible system architectures for location anonymization:
  - Centralized trusted third-party location anonymizer:
    - Such anonymization proxy server takes care of location updates and location anonymization.
  - Client-based non-cooperative location anonymization:
    - Client-based knowledge and special client-server protocols are used to maintain the client's location privacy.
  - Decentralized cooperative P2P protocols to protect privacy:
    - A Group of mobile clients collaborate with one another to provide location privacy of a single user without involving a centralized trusted authority.



Overview of Approaches



Approaches to Preserve Location Privacy

Policies

Modification of Request

Dummy Requests

Provider Change

Pseudonymization:
- Transaction pseudonym
- Role pseudonym
- Anonymization

Deliberate Impairment of Locating

Transformation/Cryptography:
- Coordinate/Space mapping
- Private Information Retrieval

Blind Out:
- Ban-Zone
- Mix-Zone

Obfuscation:
- Spatial
- Temporal

Path Confusion

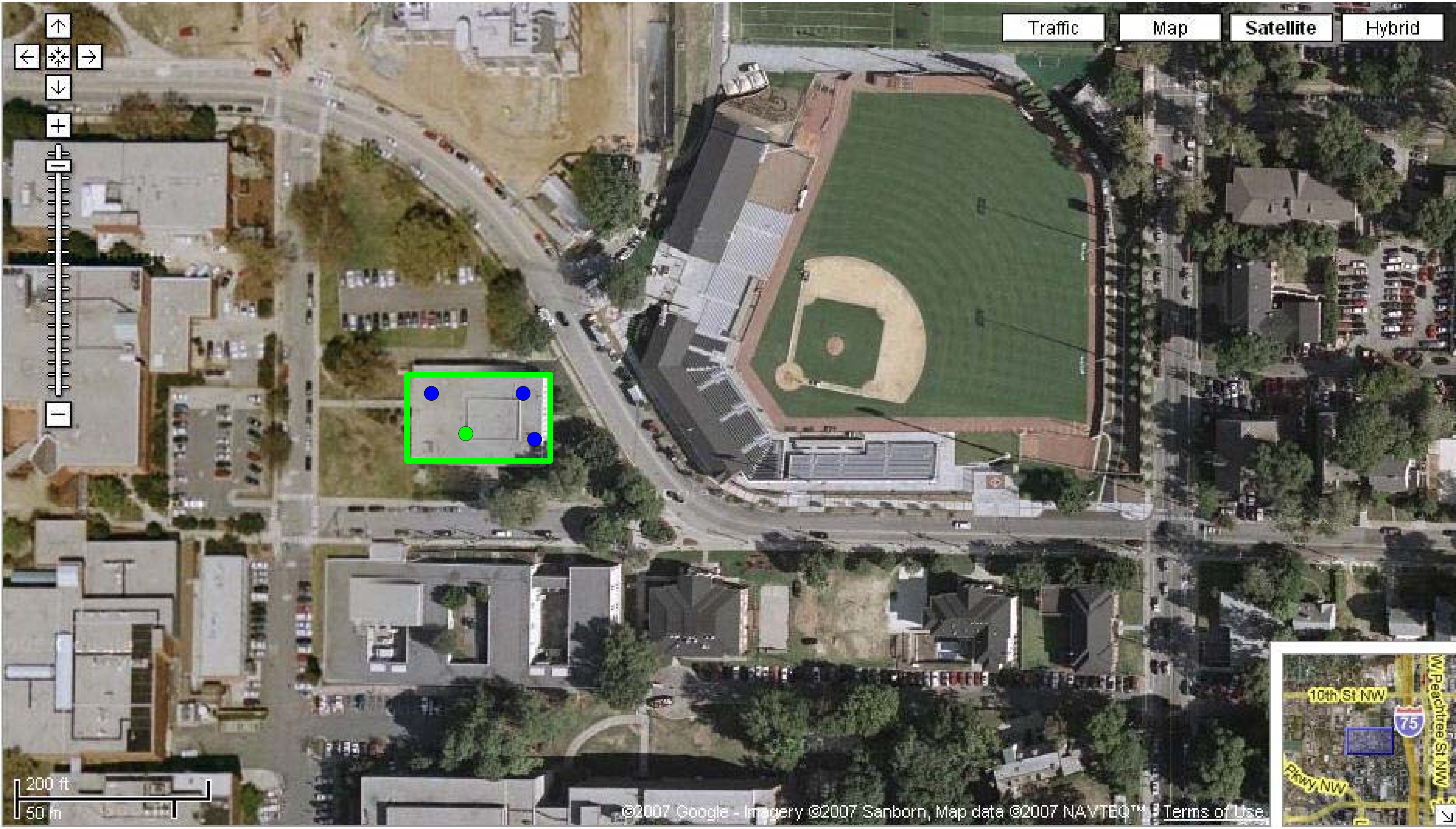
Obfuscation

- Obfuscation: deliberate reduction in precision of location
  - May be acceptable by the service:

		Spatial precision	
		High	Low
Temporal precision	Low	Turn-by-Turn On-line Navigation, POI-Finder, Tourist-Guide	Weather Notifications, Time-Critical Ads
	High	Mobile Blogging, Virtual Grqafitti/Memo, Road Hazard Detection, Mobile Data Gathering	Locatinon-Aware News, Weather Forecast

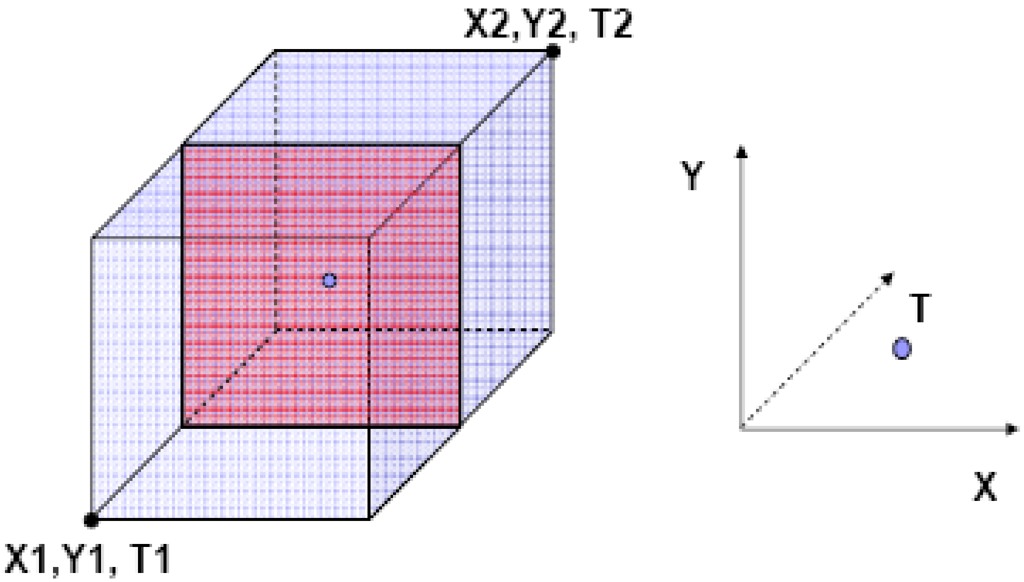
  - If not, filter-refinement approach is used:
    - The LBS server sends all the answers that are relevant to the obfuscated position
    - Anonymizer or client itself computes selects the true answer

Obfuscation: Spatial Cloaking


Traffic
Map
Satellite
Hybrid
200 ft
50 m
©2007 Google - Imagery ©2007 Sanborn, Map data ©2007 NAVTEQ™ - Terms of Use
10th St NW
W Peachtree St NW
75
Pkwy NW

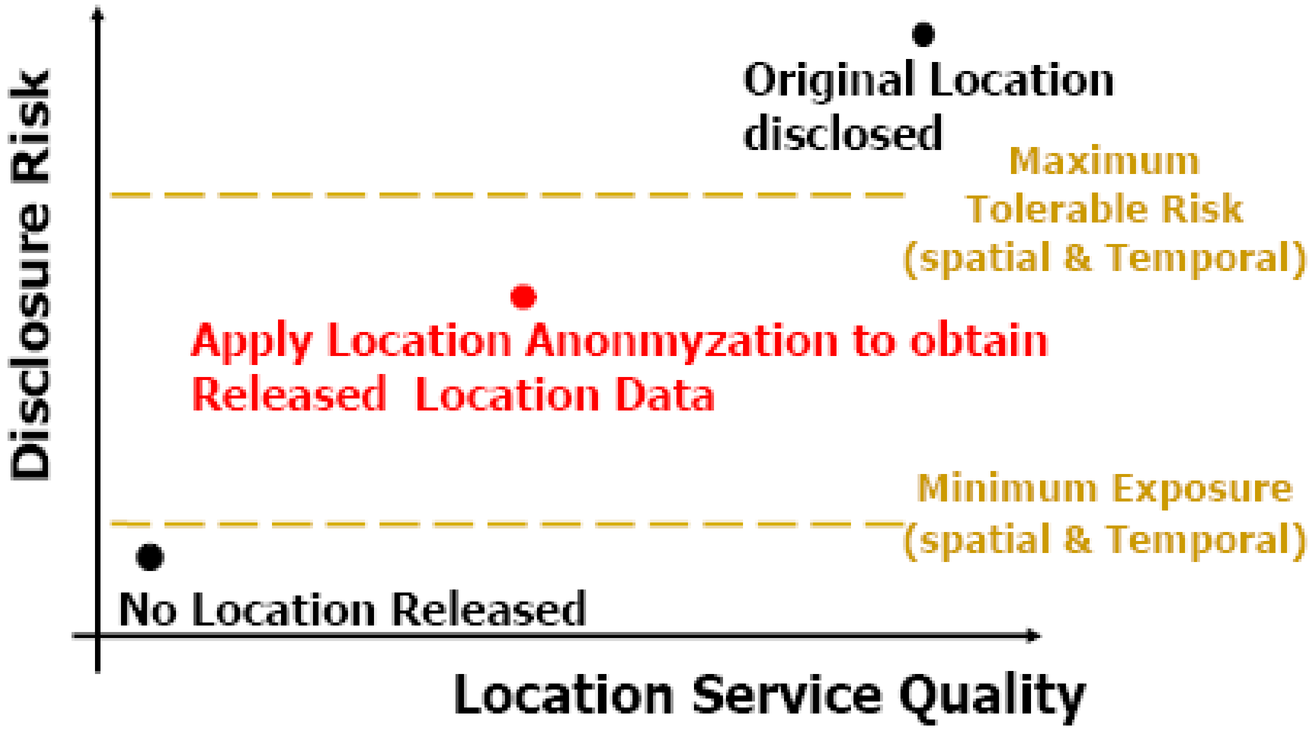
Obfuscation: Spatio-Temporal Cloak

- Spatial Cloaking first followed by temporal cloaking



How Much Cloaking: Trade-offs

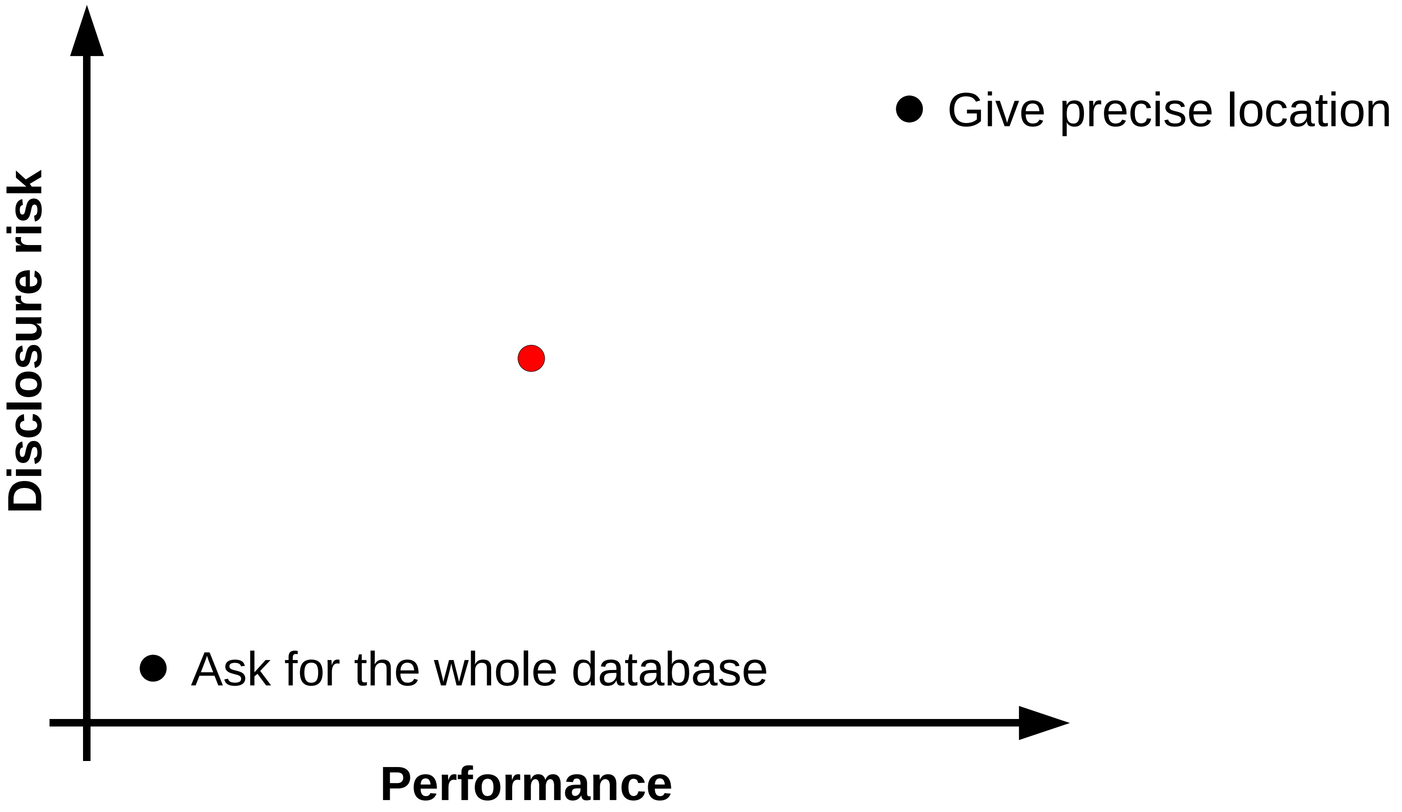
- Location privacy and LBS quality trade-off


Original Location disclosed
Maximum Tolerable Risk (spatial & Temporal)
Apply Location Anonmyzation to obtain Released Location Data
Minimum Exposure (spatial & Temporal)
No Location Released
Disclosure Risk
Location Service Quality

- [GedikLiu-ICDCS 2005, TMC 2007]

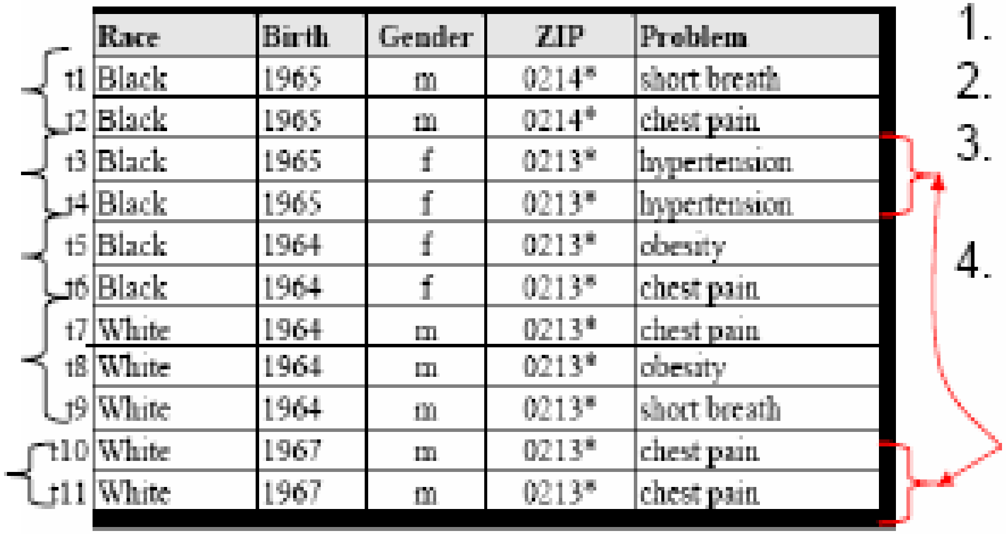
How Much Cloaking: Trade-offs

- Location privacy and LBS performance trade-off



K-Anonimity

- *How to chose the size of the cloaking region*? (ASR – anonymization spatial region)
- K-Anonimity [Samarati & Sweeney]: a concept from *privacy-preserving data mining*.
  - Goal: Preserving individual privacy while allowing public release of information
    - K-anonymity: Each tuple is indistinguishable from at least k-1 others.



	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

Example of *k*-anonymity, where *k*=2 and QI={*Race, Birth, Gender, ZIP*}

1. Identify quasi identifier
2. Remove identifier of each record
3. Ensure k-anonymity of sensitive data columns on quasi-identifier
4. Ensure l-diversity of sensitive data columns

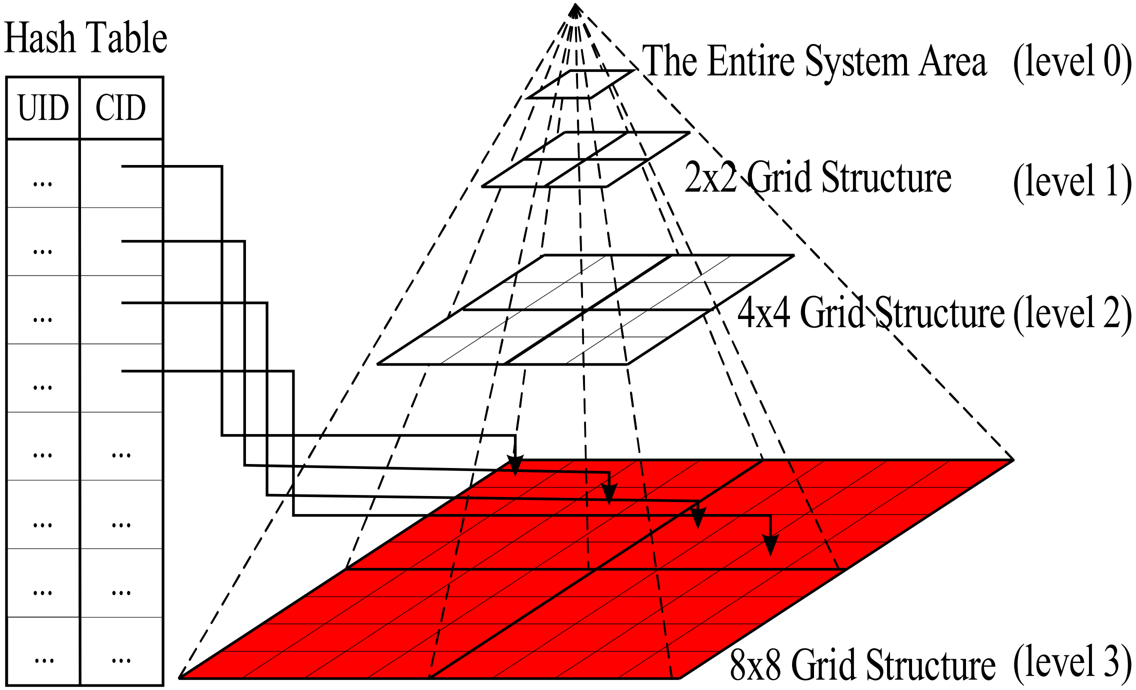
Violate l-diversity for *l* =2

Location k-Anonymity

- Location k-anonymity
  - Make sure for each location query message, there are at least k-1other messages (entries) with the same location information, each associated with a different (pseudo) identity
  - It guarantees that the adversary can only associate location information to k participants instead of to a particular individual/group/institution through inference attacks
- Location *l*-diversity (PrivacyGrid, [Bamba et al., WWW 2008])
  - For each location query message, in addition to user level k-anonymity (k different user identities), there are at least *l* different still location objects associated with each of the k users.

New Casper [Mokbel et al., VLDB 2006]

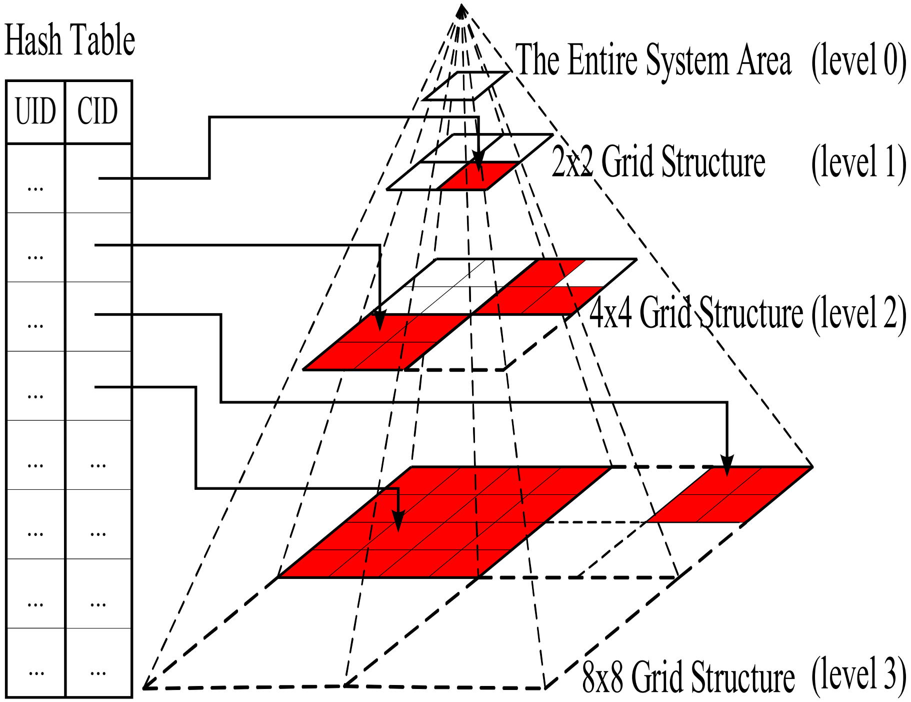
- Architecture with anonymizer
  - The entire system area is divided into grids.
  - The Location Anonymizer incrementally keeps track the number of users residing in each grid.


Hash Table
UID
CID
...
The Entire System Area (level 0)
2x2 Grid Structure (level 1)
4x4 Grid Structure (level 2)
8x8 Grid Structure (level 3)

- Traverse the pyramid structure from the bottom level to the top level, until a cell satisfying the user privacy profile is found.
- Disadvantages:
  - High location update cost.
  - High searching cost

New Casper [Mokbel et al., VLDB 2006]

- Adaptive Location Anonymizer
  - Each sub-structure may have a different depth that is adaptive to the environmental changes and user privacy requirements.



Hash Table
UID
CID
...
The Entire System Area (level 0)
2x2 Grid Structure (level 1)
4x4 Grid Structure (level 2)
8x8 Grid Structure (level 3)

- Cell Splitting: A cell *cid* at level *i* needs to be split into four cells at level *i*+1 if there is at least one user *u* in *cid* with a privacy profile that can be satisfied by some cell at level *i*+1.
- Cell Merging: Four cells at level *i* are merged into one cell at a higher level *i*-1 only if all users in the level *i* cells have strict privacy requirements that cannot be satisfied within level *i*.

Outline

- Location Privacy – An Overview
  - Assumptions, requirements, and challenges
  - Location privacy problems (attacks on privacy)
  - High-level overview of the proposed solutions

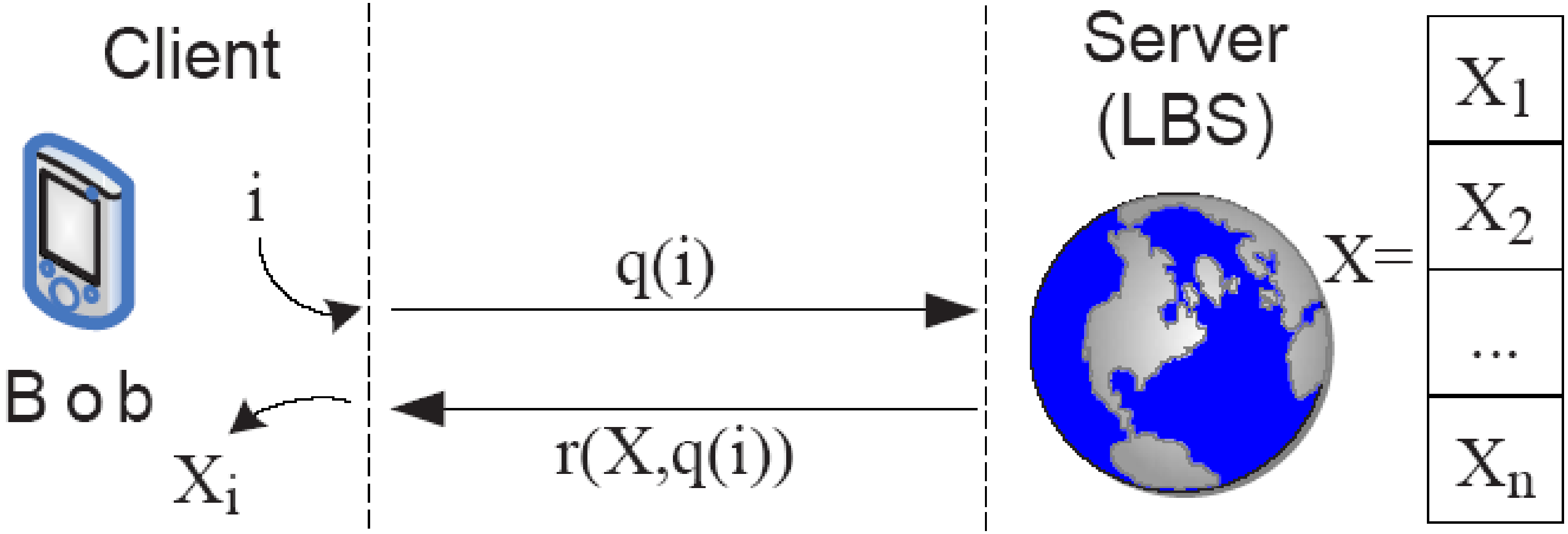
- *G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan, "Private Queries in Location-Based Services: Anonymizers are Not Necessary," ACM SIGMOD 2008***

** Acknowledging material from P.Kalnis slides

Motivation

- Limitations of existing solutions
  - Assumption of trusted entities
    - anonymizer and trusted, non-colluding users
  - Considerable overhead for sporadic benefits
    - maintenance of user locations
  - No privacy guarantees
    - especially for continuous queries

PIR Overview



- Computationally hard to find *i* from *q*(*i*)
- Bob can easily find X_i from *r* (trap-door)

PIR Theoretical foundations

- Let *N* =q_1*q_2, where q_1, and q_2 are large primes

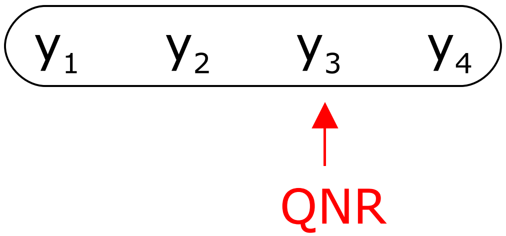
$$\mathbb{Z}_N^* = \{x \in \mathbb{Z}_N \mid gcd(N, x) = 1\}$$
$$QR = \{y \in \mathbb{Z}_N^* \mid \exists x \in \mathbb{Z}_N^* : y \equiv x^2 \bmod N\}$$

- Quadratic Residuosity Assumption (QRA)
  - QR/QNR decision computationally hard
  - Essential properties:
    - QR * QR = QR
    - QR * QNR = QNR

PIR Protocol for Binary Data

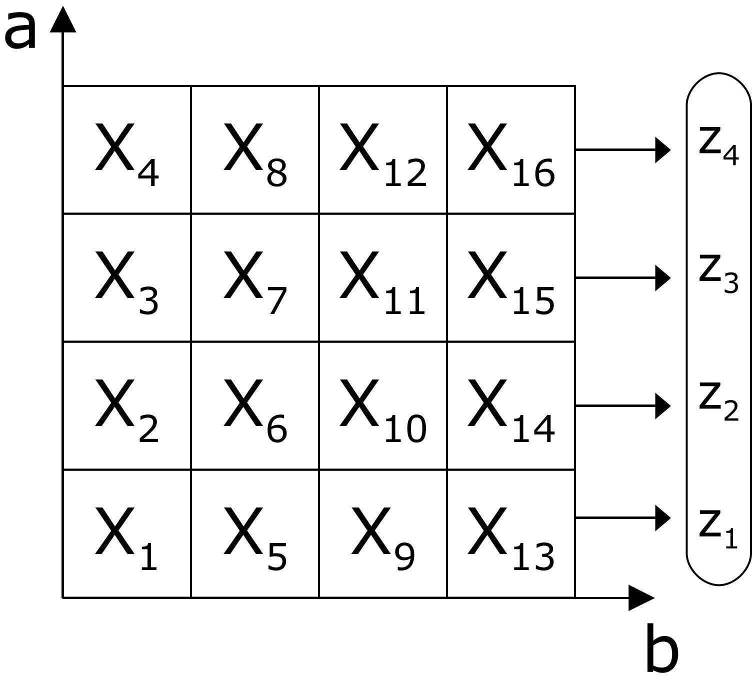
N one-bit records are organized into $\sqrt{N} \times \sqrt{N}$ matrix

u



Get X_{10}

a=2, b=3



X_4	X_8	X_{12}	X_{16}
X_3	X_7	X_{11}	X_{15}
X_2	X_6	X_{10}	X_{14}
X_1	X_5	X_9	X_{13}

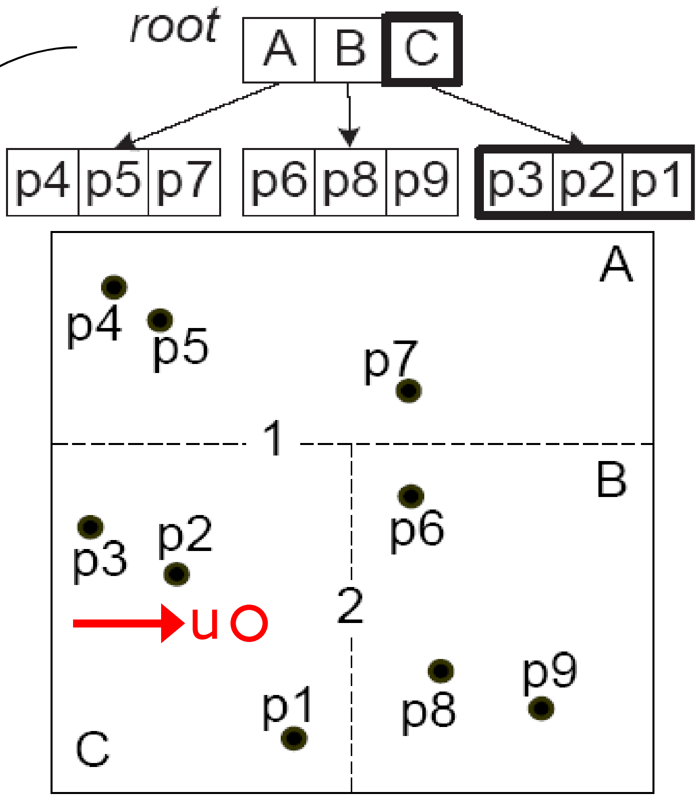
$$z_i = \prod_{j=1}^{4} X_{4\cdot(j-1)+i} \cdot y_j$$

z_2=QNR => X_{10}=1

z_2=QR => X_{10}=0

Approximate Nearest Neighbor

p_4	p_6	p_1
p_5	p_8	p_2
p_7	p_9	p_3



(a) kd-Tree

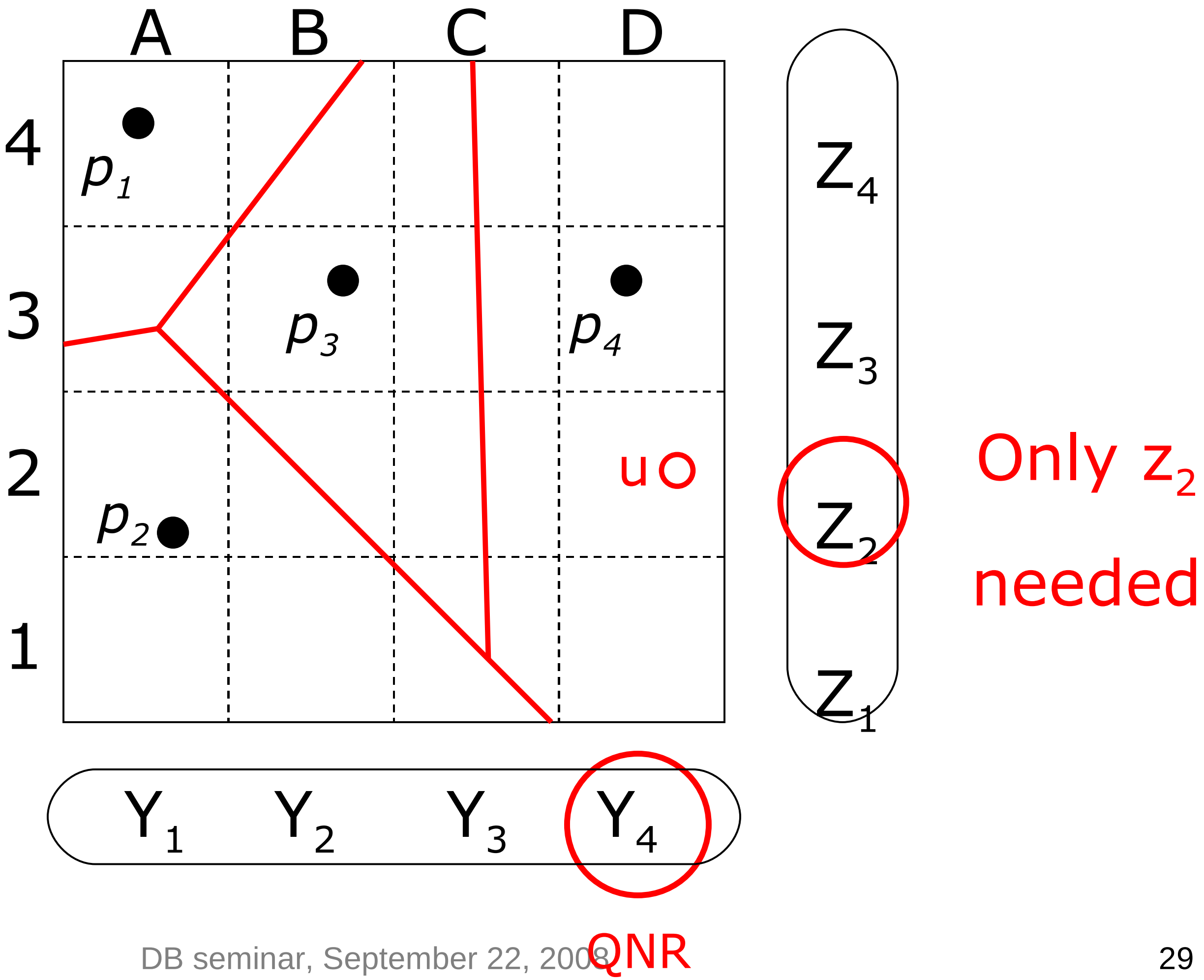
- Data organized as a square matrix
  - Each column corresponds to index leaf
  - An entire leaf is retrieved – the closest to the user

Exact Nearest Neighbor

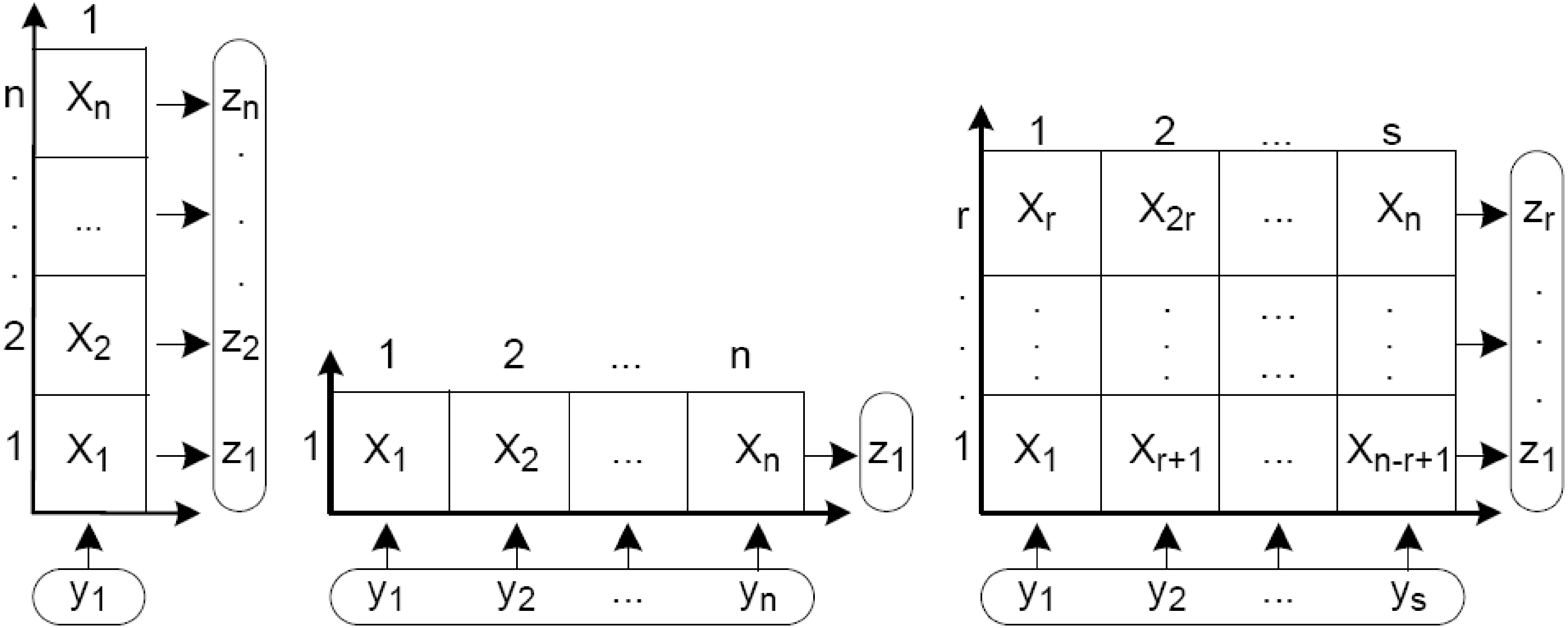
- Voronoi diagram of POIs and a regular grid is used
  - Data base size is proportional to the grid size

A3: p_1, p_2, p_3

A4: p_1, --, --



Rectangular PIR Matrix

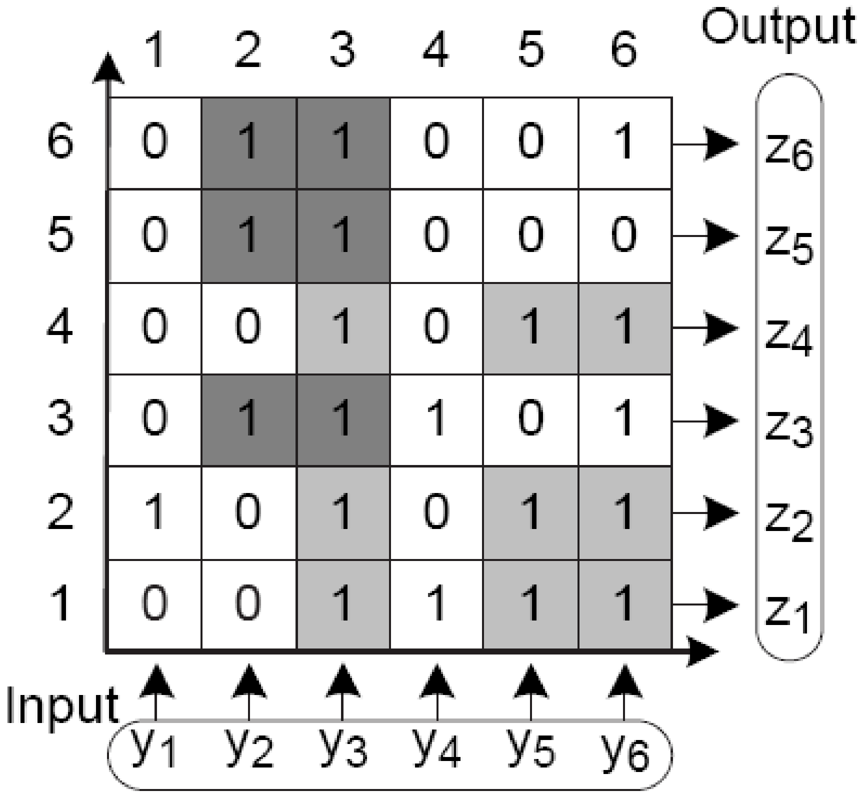


(a) M: n×1

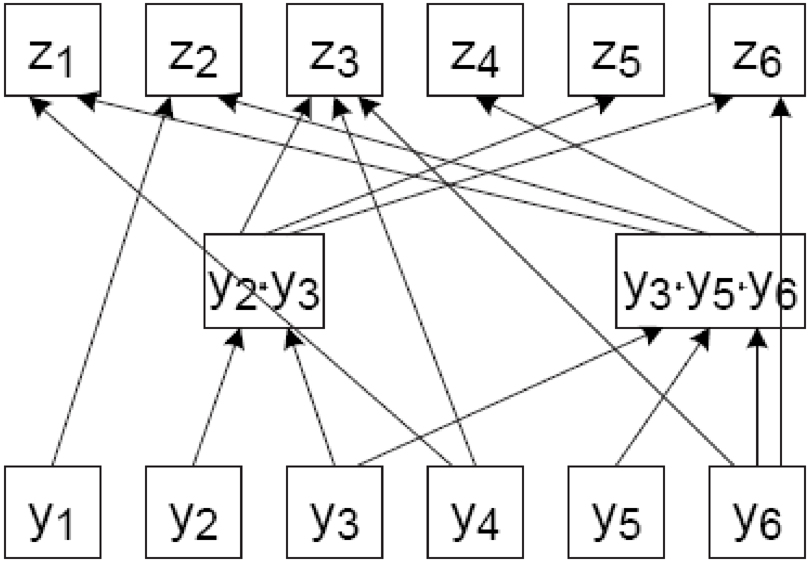
(b) M: 1×n

(c) M: r×s

Avoiding Redundant Computations



	1	2	3	4	5	6	Output
6	0	1	1	0	0	1	z_6
5	0	1	1	0	0	0	z_5
4	0	0	1	0	1	1	z_4
3	0	1	1	1	0	1	z_3
2	1	0	1	0	1	1	z_2
1	0	0	1	1	1	1	z_1
Input	y_1	y_2	y_3	y_4	y_5	y_6	



Execution plan

- Data mining
  - Identify frequent partial products

Other Optimizations

- Output from the server (*z* values) can be compressed (up to 90% in experiments), saving communication
- Values of *z* can be computed in parallel
  - Master-slave paradigm
  - Offline phase: master scatters PIR matrix
  - Online phase:
    - Master broadcasts *y*
    - Each worker computes *z* values for its strip
    - Master collects *z* results

LBS with PIR: pros/cons

- Pros:
  - Two-party cryptographic protocol
    - No trusted anonymizer required
    - No trusted users required
  - No pooling of a large user population required
    - No need for location updates
  - Location data completely obscured
- Cons:
  - Quite complex