Private Queries in Location-Based Services: Anonymizers are Not Necessary

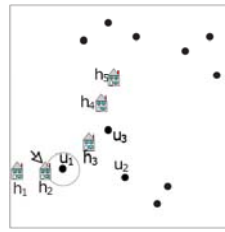
Gabriel Ghinita[1]
Panos Kalnis[1]
Ali Khoshgozaran[2]
Cyrus Shahabi[2]
Kian Lee Tan[1]

[1] National University of Singapore
[2] University of Southern California

Location-Based Services (LBS)

- LBS users
  - Mobile devices with GPS capabilities
- Queries
  - NN Queries
  - Location server is NOT trusted

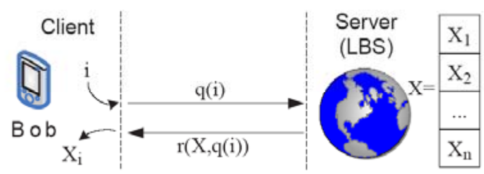
"Find closest hospital to my present location"



Problem Statement

- Queries may disclose sensitive information
  - Query through anonymous web surfing service
- But user location may disclose identity
  - Triangulation of device signal
  - Publicly available databases
  - Physical surveillance
- How to preserve *query source anonymity*?
  - Even when exact user locations are known

PIR Overview

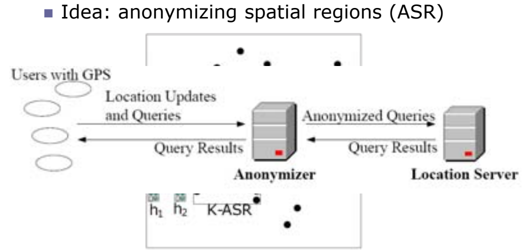


- Computationally hard to find *i* from *q(i)*
- Bob can easily find X_i from *r* (trap-door)

Existing LBS Privacy Solutions

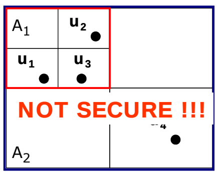
Spatial K-Anonymity

- Query issuer "hides" among other K-1 users
  - Probability of identifying query source ≤ 1/K
  - Idea: anonymizing spatial regions (ASR)



Casper[Mok06]

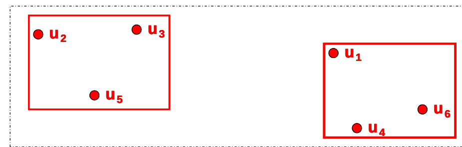
- Quad-tree based
  - Fails to preserve anonymity for outliers
  - Unnecessarily large ASR size



- Let K=3
- If any of u_1, u_2, u_3 queries, ASR is A_1
- If u_4 queries, ASR is A_2
- u_4's identity is disclosed

[Mok06] – Mokbel et al, The New Casper: Query Processing for Location Services without Compromising Privacy, VLDB 2006

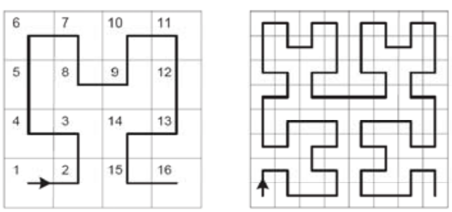
Reciprocity



[KGMP07] – Kalnis P., Ghinita G., Mouratidis K., Papadias D., "Preventing Location-Based Identity Inference in Anonymous Spatial Queries", IEEE TKDE 2007.

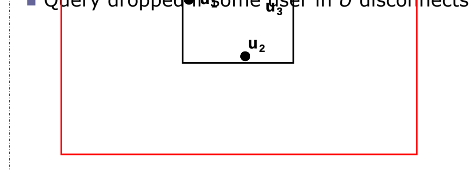
Hilbert Cloak (HC)

- Based on Hilbert space-filling curve
  - index users by Hilbert value of location
  - partition Hilbert sequence into "K-buckets"



Continuous Queries[CM07]

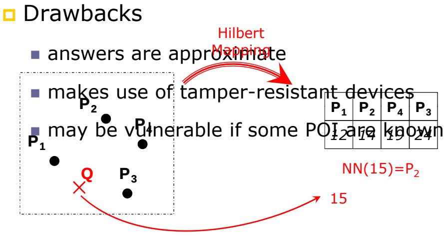
- Problems
  - ASRs grows large
  - Query dropped if some user in *U* disconnects



[CM07] C.-Y. Chow and M. Mokbel "Enabling Private Continuous Queries For Revealed User Locations". In Proc. of SSTD 2007

Space Encryption[KS07]

- Drawbacks
  - answers are approximate
  - makes use of tamper-resistant devices
  - may be vulnerable if some POI are known



[KS07] A. Khoshgozaran, C. Shahabi. Blind Evaluation of Nearest Neighbor Queries Using Space Transformation to Preserve Location Privacy , In Proc. Of SSTD 2007

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

Our Approach

LBS Privacy with PIR

- PIR
  - 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

PIR Theoretical Foundations

- Let $N = q_1 * q_2$, q_1 and q_2 large primes

$$\mathbb{Z}_N^* = \{x \in \mathbb{Z}_N | gcd(N, x) = 1\}$$

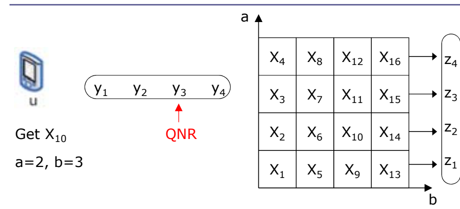
$$QR = \{y \in \mathbb{Z}_N^* | \exists x \in \mathbb{Z}_N^* : y = x^2 \mod N\}$$

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

*QR * QR = QR*

*QR * QNR = QNR*

PIR Protocol for Binary Data

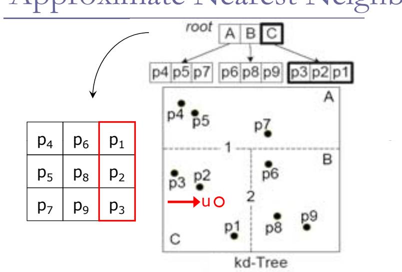


$$z_i = \prod_{j=1}^{4} X_{4\cdot(j-1)+i} \cdot y_j$$

z_2=QNR => X_{10}=1

z_2=QR => X_{10}=0

Approximate Nearest Neighbor

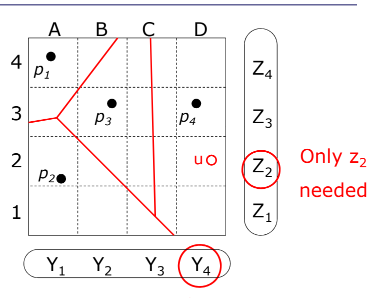


- Data organized as a square matrix
  - Each column corresponds to index leaf
  - An entire leaf is retrieved – the closest to the user

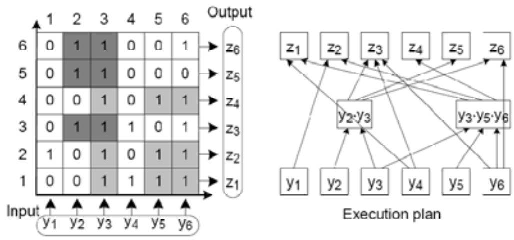
Exact Nearest Neighbor

A3: p_1, p_2, p_3

A4: p_1, --, --



Avoiding Redundant Computations



- Data mining
  - Identify frequent partial products

Parallelize Computation

- 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

Experimental Settings

- Sequoia dataset + synthetic sets
  - 10,000 to 100,000 POI
- Modulus up to 1280 bits

Computation/Communication Overhead (Approximate)

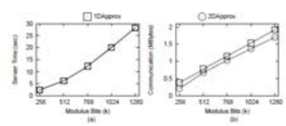


Figure 16: Variable k, Sequoia set (62K POI)

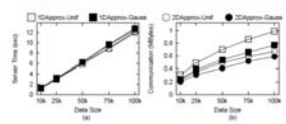


Figure 17: Variable data size, $k = 768$ bits

Computation/Communication Overhead (Exact)

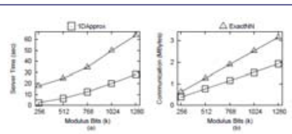


Figure 19: Variable k, Sequoia set (62K POI)

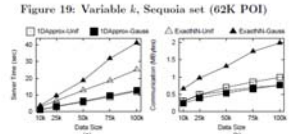
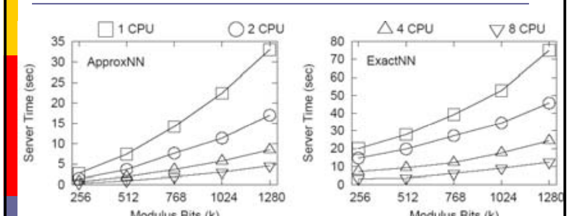
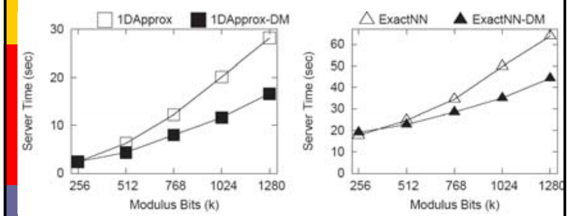


Figure 20: Variable data size, $k = 768$ bits

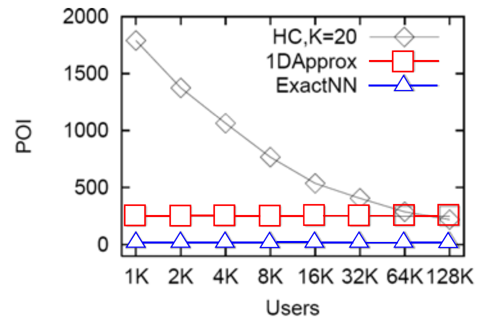
Parallel Execution



Data Mining Optimization



Disclosed POI



Conclusions

- PIR-based LBS privacy
  - No need to trust third-party
  - Secure against any location-based attack
- Future work
  - Further reduce PIR overhead
  - Support more complex queries
  - Include more POI information in the reply

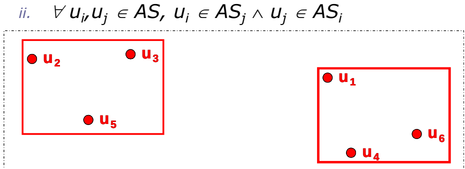
Discussion

- Given the parallelization, compression, multiplication reduction, rectangular shape M, how much is communication/computation saved?
- How do you compare the previous two approaches?
- What do *you* think is the major challenge in achieving privacy-aware LBS?

Privacy ⟷ Efficiency

Reciprocity

- Consider querying user u_q and ASR A_q
- Let AS_q = {set of users enclosed by A_q}
- A_q has the reciprocity property iff
  i. $|AS| \geq K$
  ii. $\forall u_i, u_j \in AS,\ u_i \in AS_j \wedge u_j \in AS_i$



[KGMP07] – Kalnis P., Ghinita G., Mouratidis K., Papadias D., "Preventing Location-Based Identity Inference in Anonymous Spatial Queries", IEEE TKDE 2007.

Continuous Queries[CM07]

- Extends reciprocity to moving clients
  - Let A_0 be ASR at time t_0, let U be the users in A_0
  - At time t_i, ASR is MBR of U (at new locations)
- Problems
  - ASR grows large
  - Query dropped if some user in U disconnects

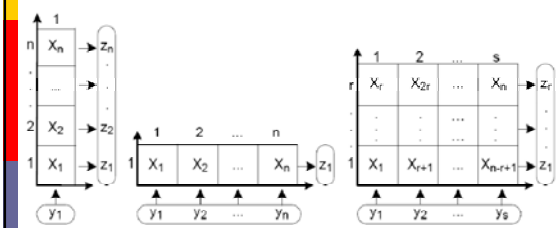
[CM07] C.-Y. Chow and M. Mokbel "Enabling Private Continuous Queries For Revealed User Locations". In Proc. of SSTD 2007

Space Encryption[KS07]

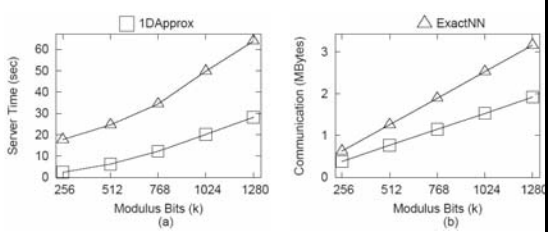
- Does not employ SKA
  - each POI is mapped to 1-D value (Hilbert)
    - fractal parameters are kept secret
  - answers are approximate
  - makes use of tamper-resistant devices
  - may be vulnerable if some POI are known

[KS07] A. Khoshgozaran, C. Shahabi. Blind Evaluation of Nearest Neighbor Queries Using Space Transformation to Preserve Location Privacy , In Proc. Of SSTD 2007

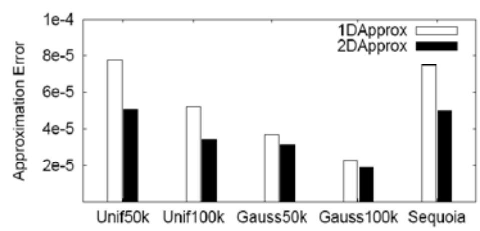
Rectangular PIR Matrix



Server Computation Overhead



Approximation Error



Bibliography

- [KGMP07] – Kalnis P., Ghinita G., Mouratidis K., Papadias D., "Preventing Location-Based Identity Inference in Anonymous Spatial Queries", IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE), 19(12), 1719-1733, 2007.
- [GZPK07] – Ghinita G., Zhao K., Papadias D., Kalnis P., *Reciprocal Framework for Spatial K-Anonymity,* Technical Report
- [GKS07a] – Ghinita G., Kalnis P., Skiadopoulos S., "PRIVE: Anonymous Location-based Queries in Distributed Mobile Systems", Proc. of World Wide Web Conf. (WWW), Banff, Canada, 371-380, 2007.
- [GKS07b] – Ghinita G., Kalnis P., Skiadopoulos S., "MOBIHIDE: A Mobile Peer-to-Peer System for Anonymous Location-Based Queries", Proc. of the Int. Symposium in Spatial and Temporal Databases (SSTD), Boston, MA, 221-238, 2007.

http://anonym.comp.nus.edu.sg