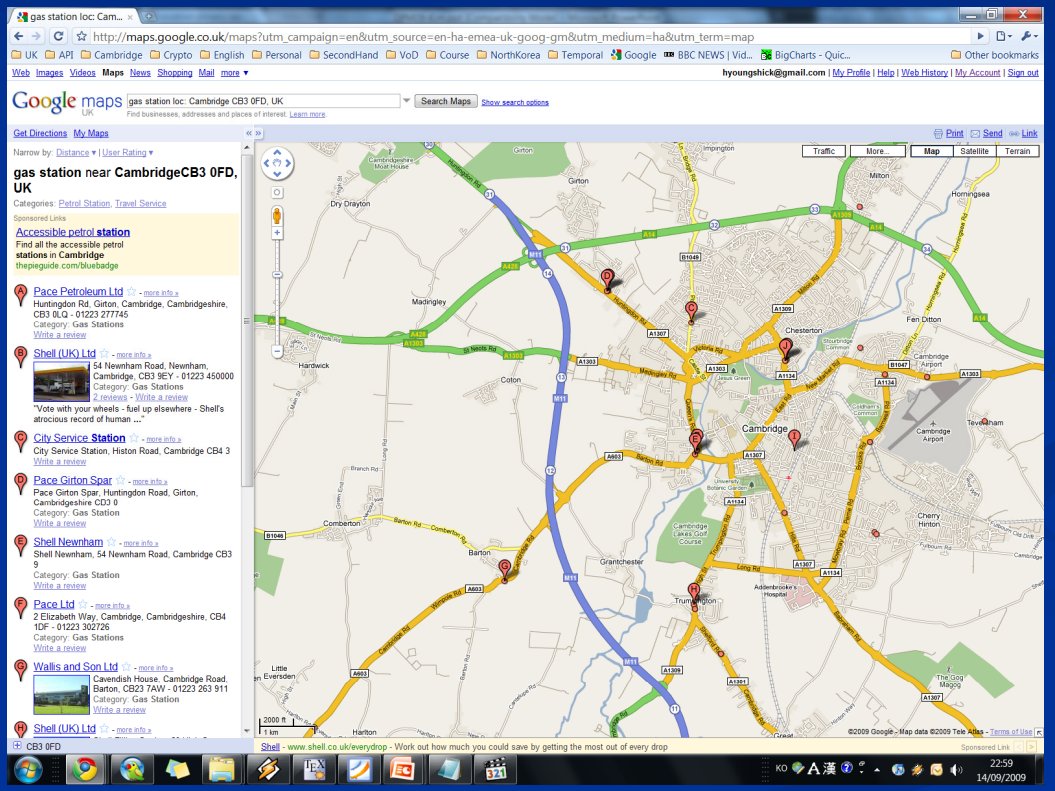
A Spatial Cloaking Framework based on Range Search for Nearest Neighbor Search

Hyoungshick Kim
Computer Laboratory
University of Cambridge

Nearest Neighbor Query

Where is the nearest POI (e.g. gas station) ?



Query Example: “gas station loc: Cambridge CB3 0FD, UK”

Query Privacy

1: Here is “Cambridge CB3 0FD, UK”

2: The nearest gas station is …

User

I do not want to give this information.

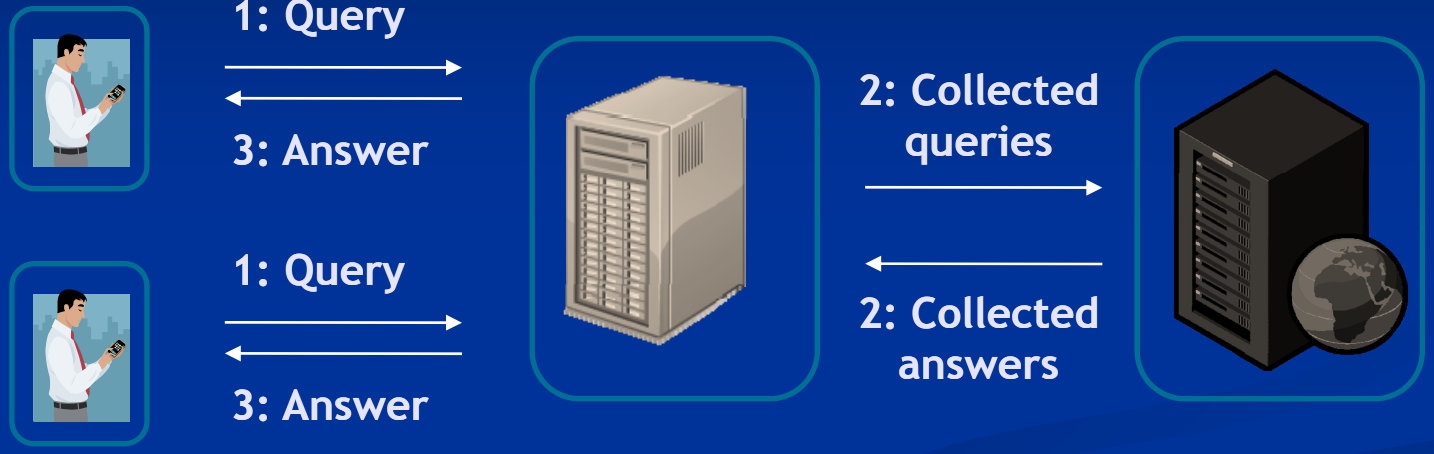
Service Provider

collects the following information about user:

- User account - physical location
- User device’s network address - physical location

In this setting, we assume the service provider is the adversary.

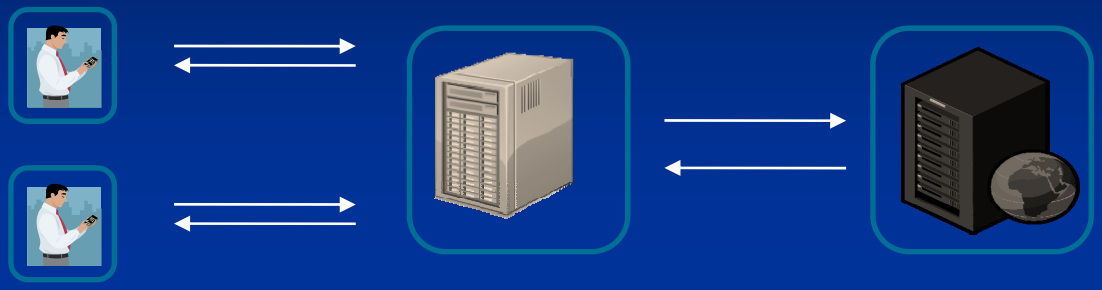
How? Use Third Party Anonymizer



- hides the relationship between queries and queriers.

Most existing systems [GG03, BF04, MCA06, BL08] are designed under the assumption of trusted anonymizers.

Limitations of Trusted Anonymizer



- Major redesign of technologies (e.g., protocols or trusted mechanism) or business models
- Single server failure/overhead
- A large number of users

Alternatives - User Centric

1: "Transformed query"

2: Answer for "Transformed query"

User

Service Provider

3: Find the nearest neighbor from the answer for "transformed query".

I cannot infer the user location from this "transformed query".

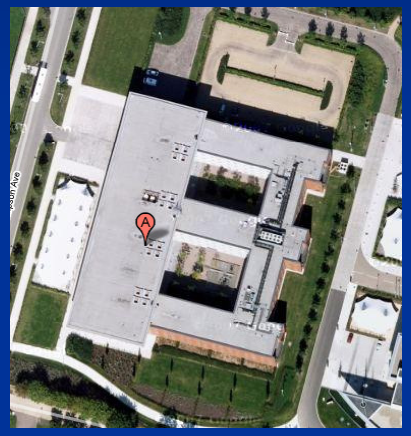
Previous Work

- False dummies [KYS05]
  - High communication/processing cost
- Transformation based on obfuscated map [KS07]
  - Approximate answer
  - A third party is still required to create an obfuscated map
- Transformation based on Private Information Retrieval (PIR) [GKKST08]
  - Theoretically secure
  - High communication/processing cost
- Incremental spatial cloaking with a fake dummy [YJHL08]
  - Incremental fetching POIs* from the service provider with a fake dummy until the user can produce the exact result
  - Multiple message rounds to stop the incremental search
  - The user's desired level of privacy (or region) cannot be guaranteed.

*POI: Point of Interest

Our Transformation

Control the granularity of location query.



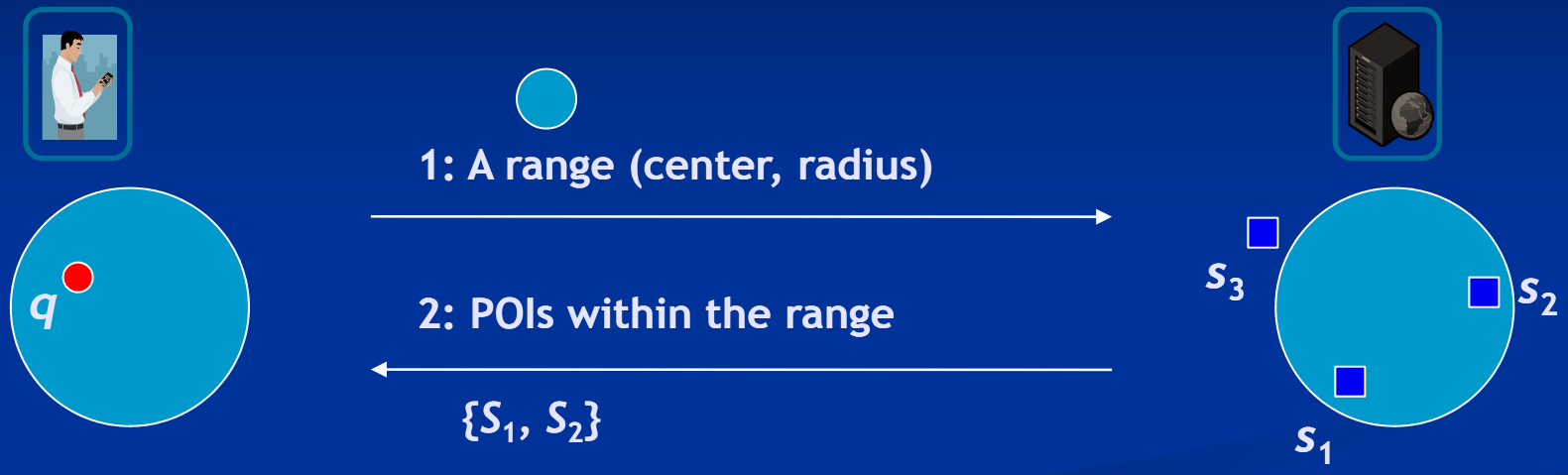
Cambridge CB3 0FD, UK

West Cambridge

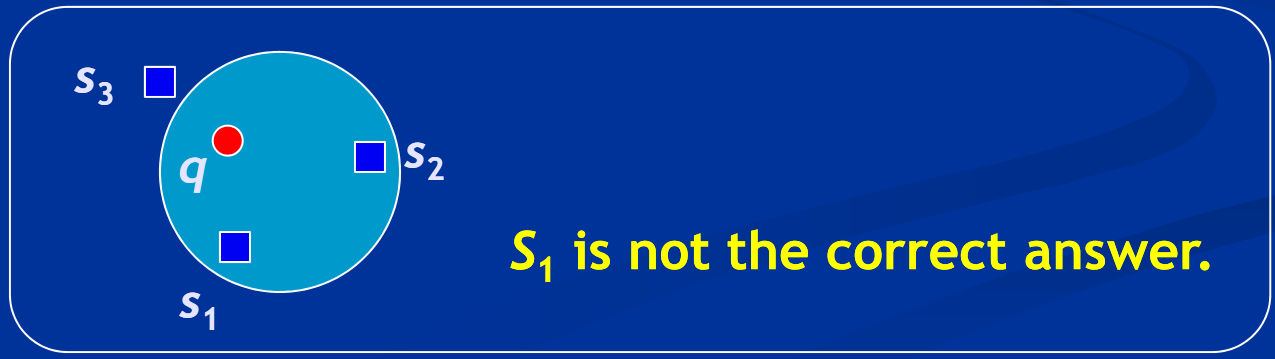
Previously, this approach seems not desirable.

- High communication cost is required.
  - But, communication cost is dramatically decreasing.
- Local search in user device is required.
  - But, computing capability of mobile devices is improving.

Naïve Range Search Query

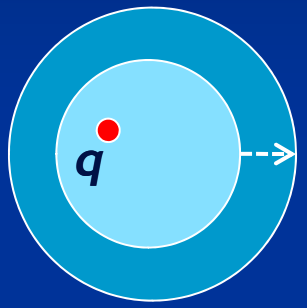


3: Choose the nearest neighbor S_1.



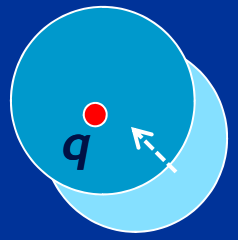
How Can We Prevent It?

1. Increase the size of range window.



- Communication cost is increasing depending on the size of window.
- A user cannot determine the optimal window size to guarantee the nearest neighbor.

2. Create the range window to locate *q* near the center of the window.



- This technique may give the information about the position *q*.

Our Approach

Our challenging issues are

- how to find the optimal range window.

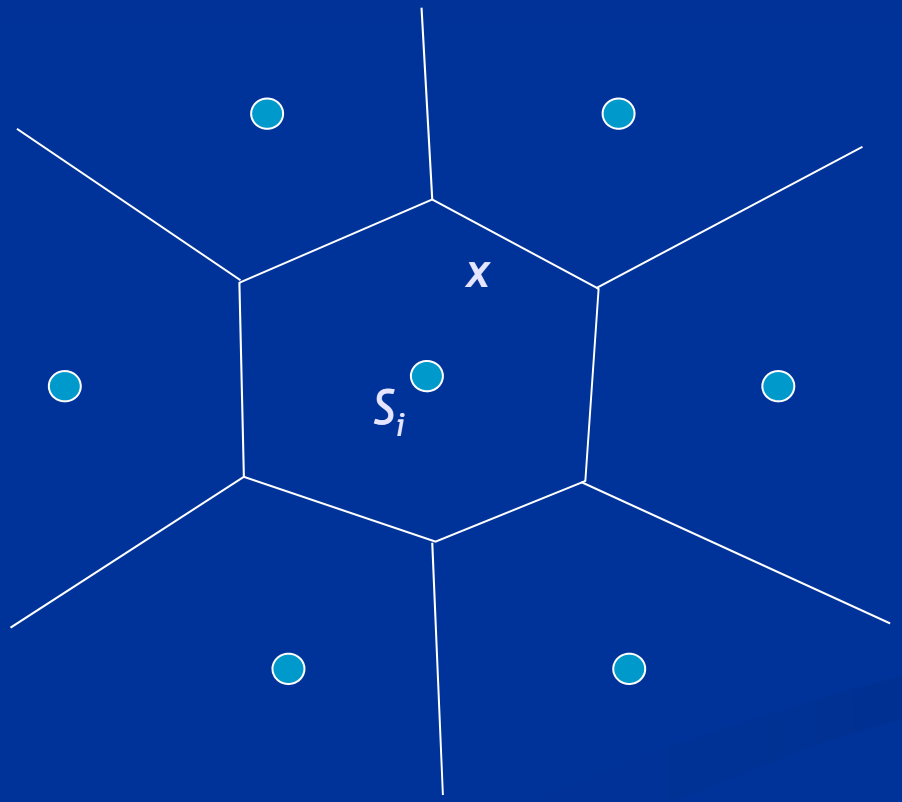
  Use the local Voronoi diagram

- how to guarantee that the user can be uniformly located at any position within the window.

  Use the fake (random) query position

Voronoi Diagram

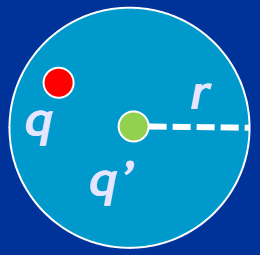
- Subdivision of plane (space) into cells
  - $S = \{S_1, S_2, \ldots S_n\}$ points in the plane
  - $V(S_i) = \{ x : \ d(x, S_i) < d(x, S_j)$ for all $j \neq i\}$



The position *x*'s the nearest neighbor is S_i.

Proposed Framework

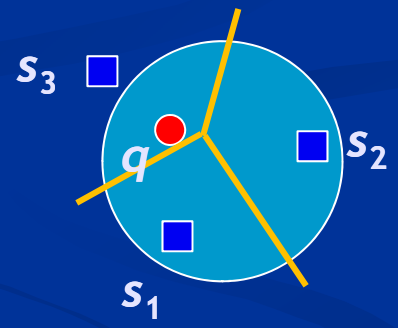
1: Given a security parameter r, generate a random circle including q with the radius r.



2: Random circle (q', r)

4: $\{S_1, S_2, S_3\}$

3: Compute the intersected Voronoi cells.

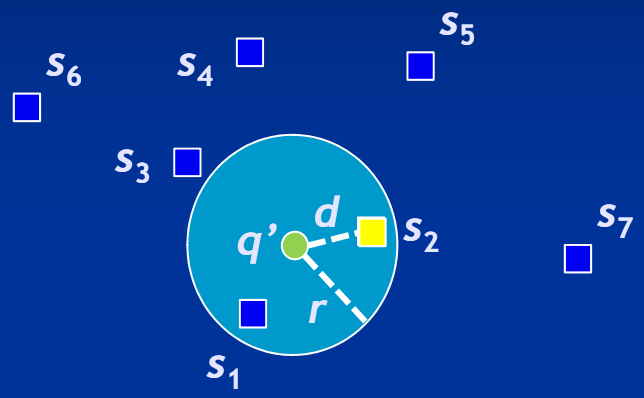


5: Choose the nearest neighbor S_3.

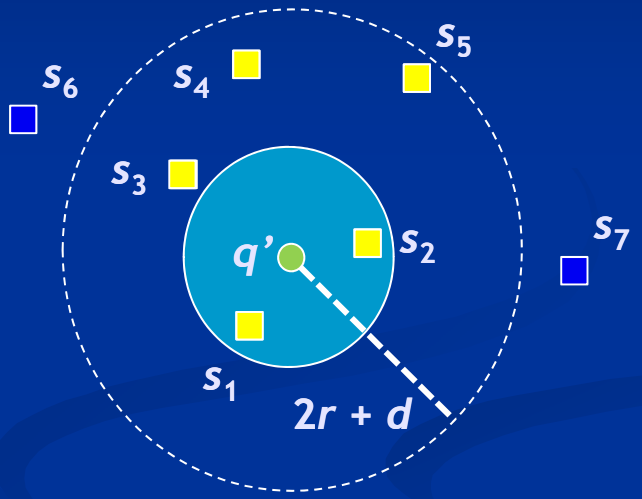
The adversary cannot obtain the information about q except that it is located with the circle.

Computation of Local Voronoi Cells

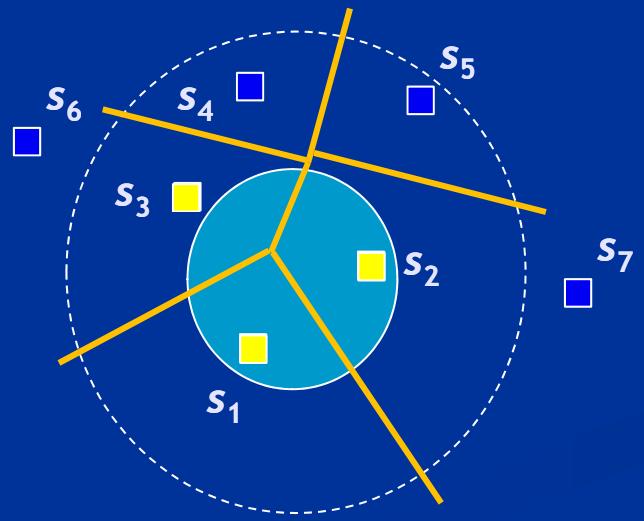
1: Find the nearest S_i from q'.



2: Find the POIs within the distance $2r + d$.



3: Find the intersected Voronoi cells.



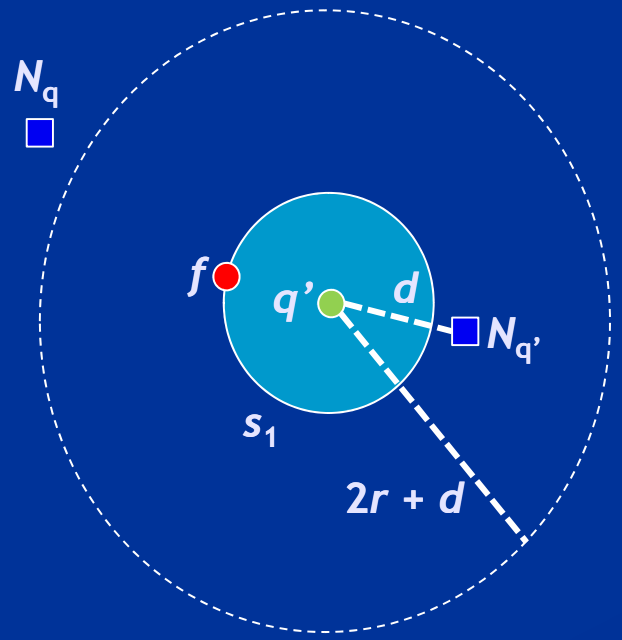
Running time (n: # of S_i, t: # of POIs within $2r+d$)

- $O(n + t \log t)$
- $O(\log n + t \log t)$ with pre-processing

Correctness of the Computation

The nearest POI (N_q) of the query position q is necessarily included in the POIs within $2r + d$ in the step 2.

Proof. Assume that N_q is not included the POIs in the step 2. From the assumption, $\text{dist}(N_q, q') \geq 2r + d$. Let f be the farthest point on the circle from $N_{q'}$.

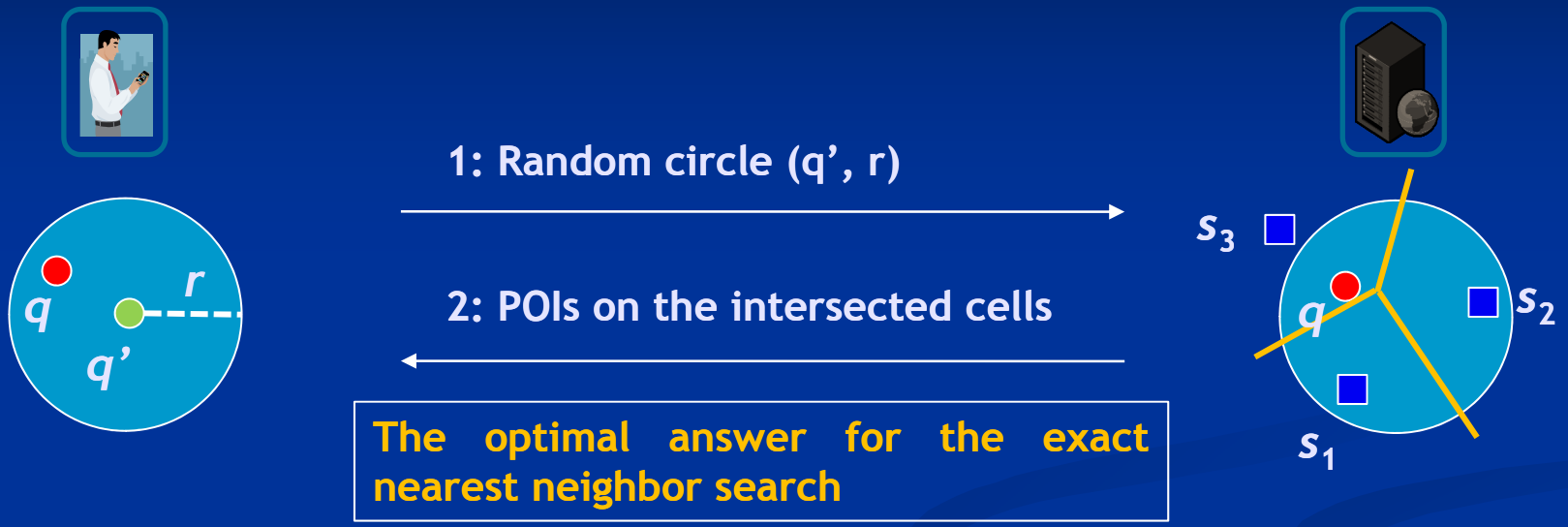


$$\text{dist}(q, N_{q'}) \leq \text{dist}(f, N_{q'}) \leq r + d \leq \text{dist}(q, N_q)$$

Therefore N_q is not the nearest POI from q.

By the contradiction, the assumption is wrong.

Inherent Problem of Range Search

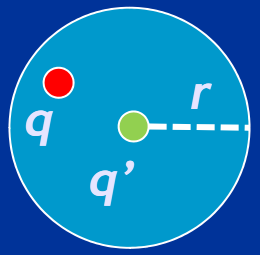


The optimal answer for the exact nearest neighbor search

However, it still requires high communication cost when a user needs a high level privacy.

Approximation

1: Given a security parameter r, generate a random circle including q with the radius r.

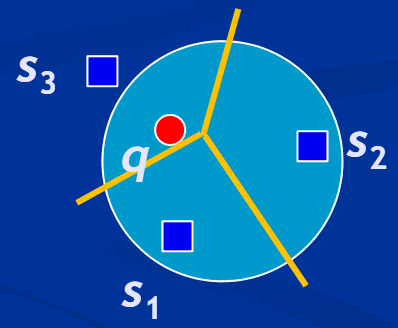


2: Random circle (q', r), Answer size k (=2)

5: $\{S_1, S_2\}$

6: Choose the nearest neighbor S_1.

3: Compute the intersected Voronoi cells.



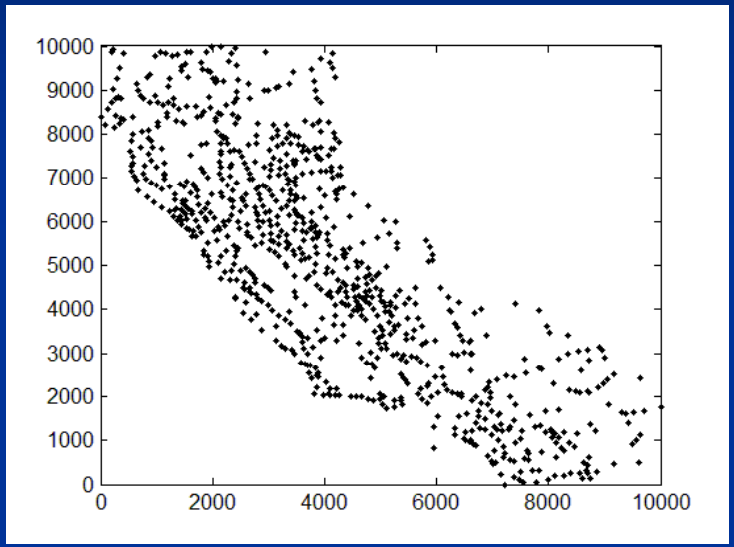
4: Select k POIs with the probability p.

$$p = \frac{\text{the intersected area of } S_i}{\text{the area of the circle}}$$

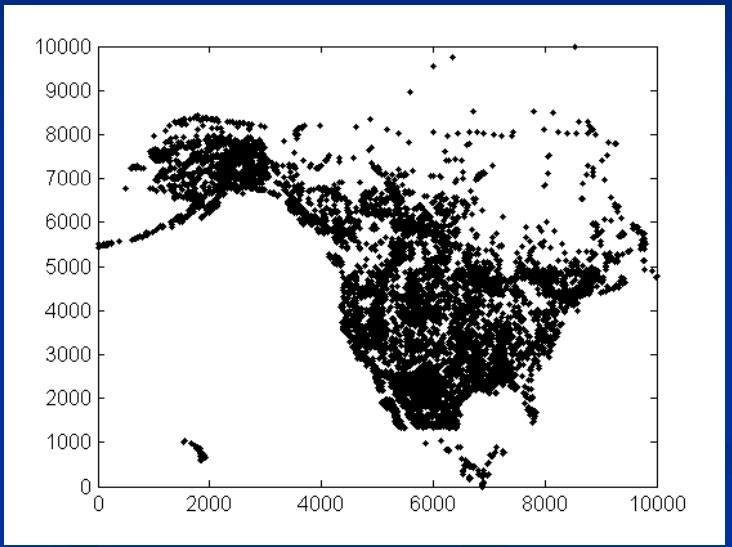
Experimental Results

Datasets

CA



NA



864 POIs

$r = 50 \backsim 1{,}550$

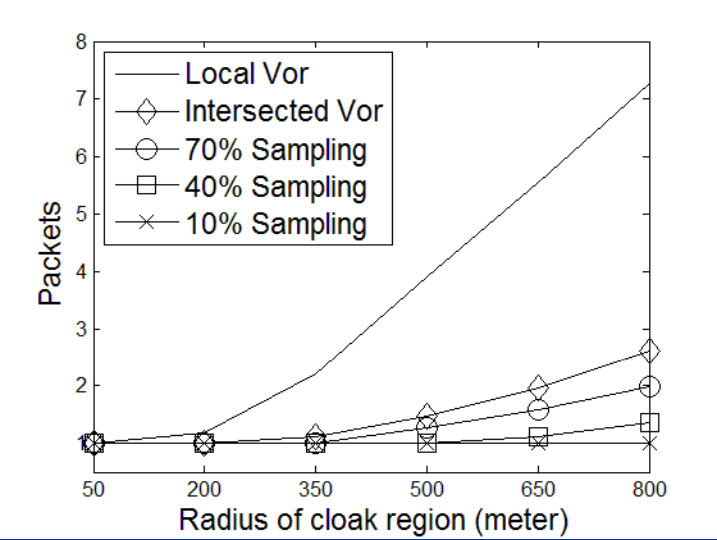
9,203 POIs

$r = 50 \backsim 1{,}050$

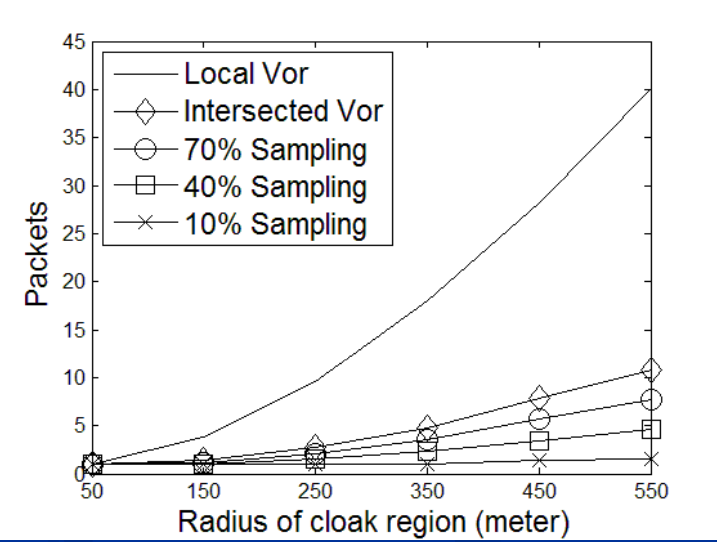
We generated 100 random queries using the Gaussian distribution of the POIs in each dataset.

Communication Cost

CA



NA



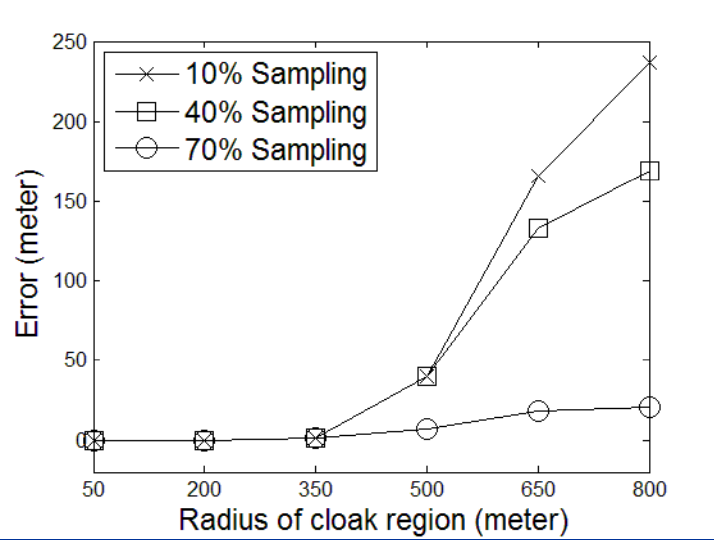
The communication cost is the number of (TCP/IP) packets transmitted.

We observe that # of packets are under 3 for CA (or 12 for NA).

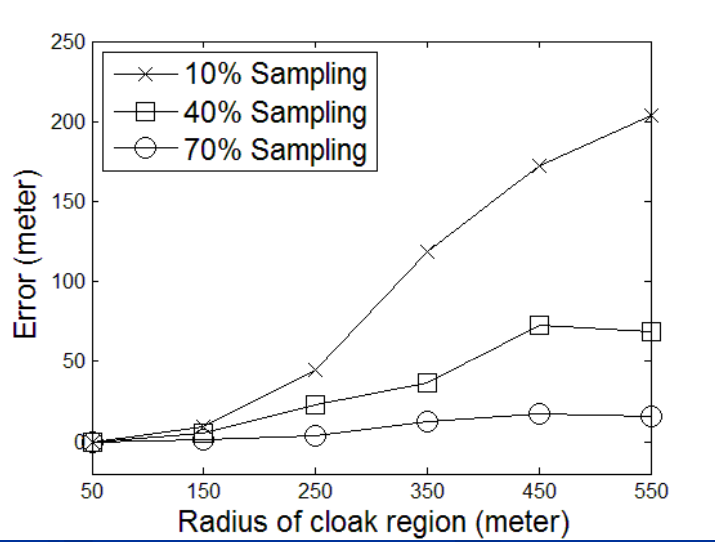
*POI size: 8 bytes, Packet header: 40-bytes, MTU: 576 bytes

Error Distance in Approximation

CA

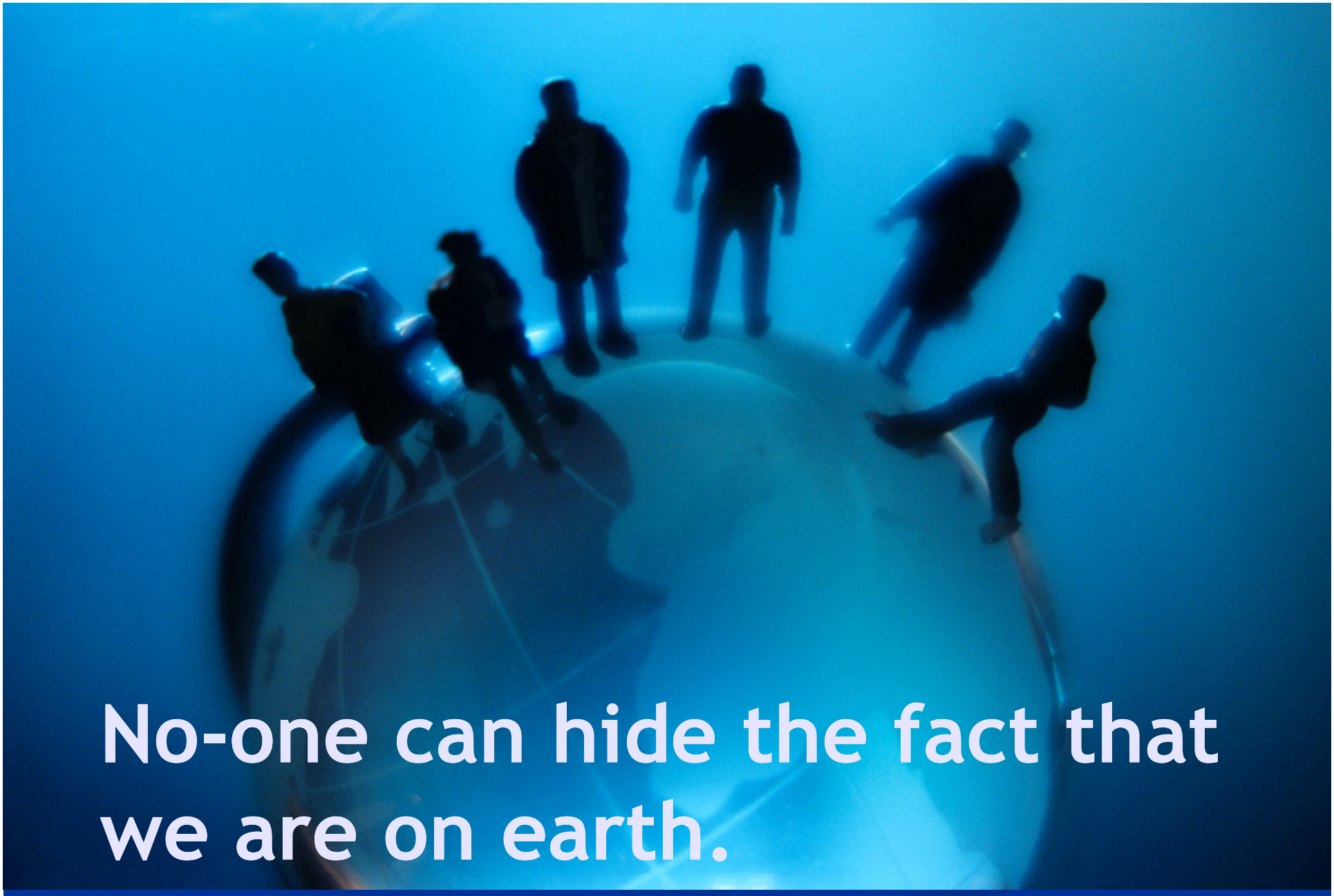


NA



- All samplings provide reasonable error distance for small *r*.
- The 70% sampling is scalable even for large *r*.

Conclusion



No-one can hide the fact that we are on earth.

Conclusion

- We show a spatial cloaking based on range search is practically enough for nearest neighbor search
  - Minimum location information leaking on range
  - Reasonable processing and communication cost due to the local Voronoi diagram
- Advantages
  - Simple client-server architecture
  - Flexible privacy level
- Future work
  - Extension to “road networks”
  - Optimal route planning

Thank you!

hk331@cl.cam.ac.uk

Related Work

- [GG03] Marco Gruteser and Dirk Grunwald. “Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking.” MobiSys 2003
- [MCA06] Mohamed F. Mokbel, Chi-Yin Chow and Walid G. Aref. “The New Casper: Query Processing for Location Services without Compromising Privacy.” VLDB 2006
- [BL08] B. Gedik and Ling Liu. “Protecting Location Privacy with Personalized k-Anonymity: Architecture and Algorithms.” IEEE Transactions on Mobile Computing In Mobile Computing 2008
- [YJHL08] Man Lung Yiu, Christian S. Jensen, Xuegang Huang and Hua Lu. “SpaceTwist: Managing the Trade-Offs Among Location Privacy, Query Performance, and Query Accuracy in Mobile Services.” ICDE 2008
- [GKKST08] Gabriel Ghinita, Panos Kalnis, Ali Khoshgozaran, Cyrus Shahabi and Kian-Lee Tan. “Private queries in location based services: anonymizers are not necessary.” SIGMOD 2008
- [BF04] Alastair R. Beresford and Frank Stajano. “Mix-zones: User privacy in location-aware services.” PerSec 2004
- [KYS05] H. Kido, Y. Yanagisawa and T. Satoh. “An anonymous communication technique using dummies for location-based services.” ICPS 2005
- [KS07] A. Khoshgozaran and C. Shahabi. “Blind Evaluation of Nearest Neighbor Queries Using Space Transformation to Preserve Location Privacy.” SSTD 2007