Privacy-Preserving Publication of User Locations in the Proximity of Sensitive Sites

Bharath Krishnamachari

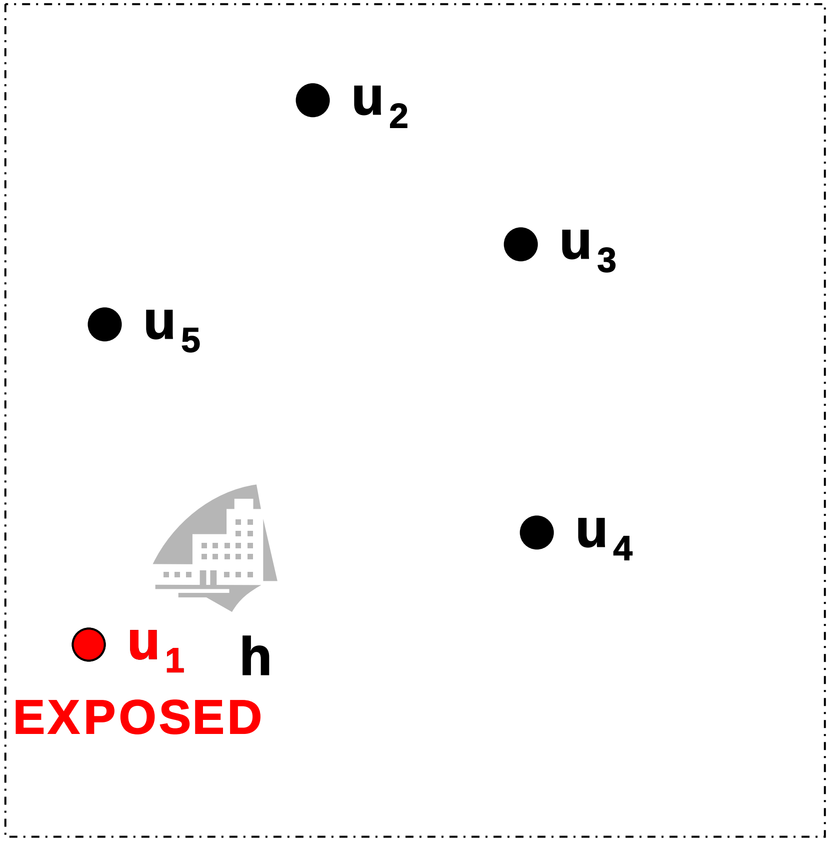
Gabriel Ghinita

Panos Kalnis

National University of Singapore

Privacy Threat

Hu et al., “Privacy Aware Location Publishing for Moving Clients”



Location Publishing

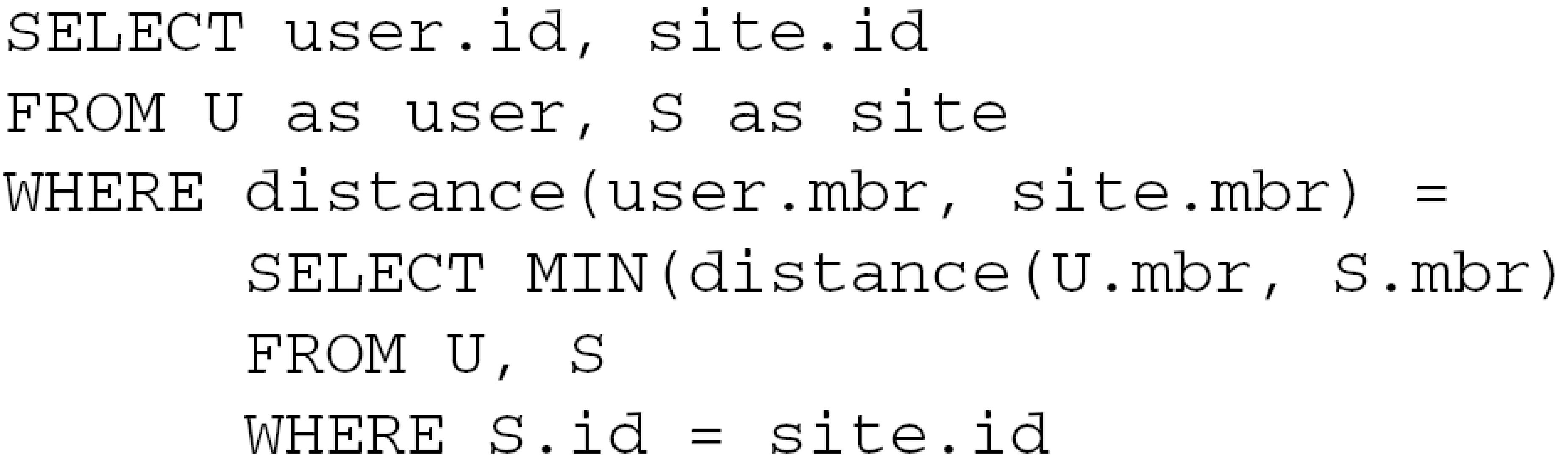
- Traffic congestion control
- Infrastructure planning
- Snapshot at 2pm

Hospital Data (external)

Time	*Specialty*
01:00 PM	Dentistry
02:00 PM	Cardiology
04:00 PM	Surgery

Attack

- Attack: Associate a site ***s*** with fewer than **K** users

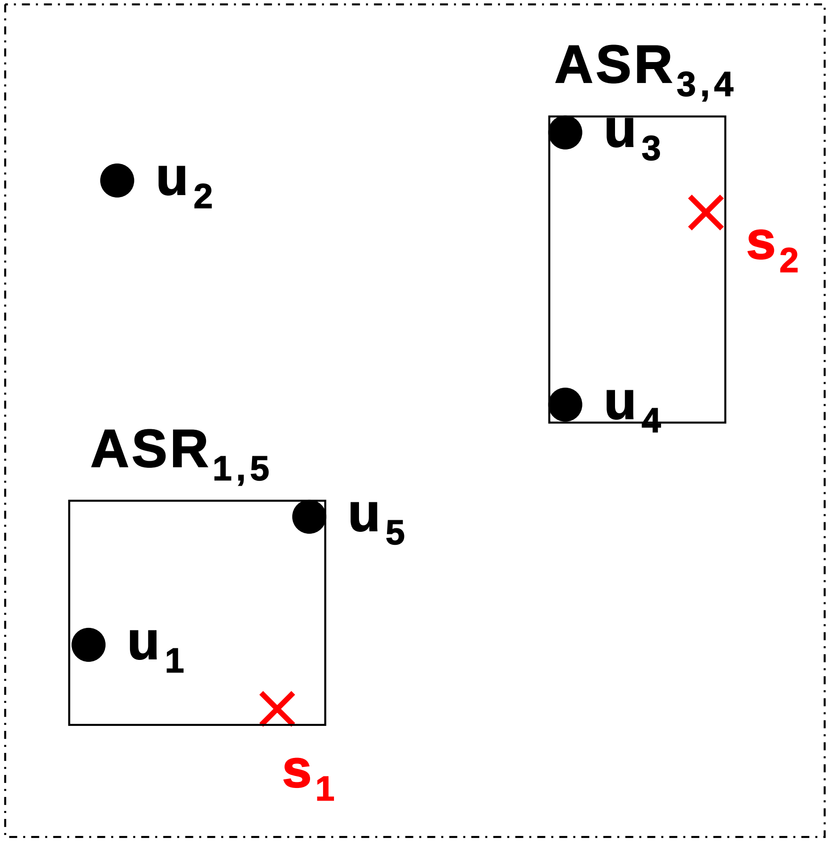


```
SELECT user.id, site.id
FROM U as user, S as site
WHERE distance(user.mbr, site.mbr) =
      SELECT MIN(distance(U.mbr, S.mbr)
      FROM U, S
      WHERE S.id = site.id
```

- Attack is successful in the previous slide:

< u_1, h >

Solution Outline (*K*-anonymity)



- Attack, K=2

<u1, s1>

<u5, s1>

<u3, s2>

<u4, s2>

NOT successful

Problem Statement

- User set *U*, sensitive sites set *S*
- Mapping *M:S -> 2^U*

$$\forall s \in S, |M(s)| = K$$

$$\forall s_1, s_2 \in S, M(s_1) \cap M(s_2) = \varnothing$$

- Minimize Generalization Cost **(GGC)**

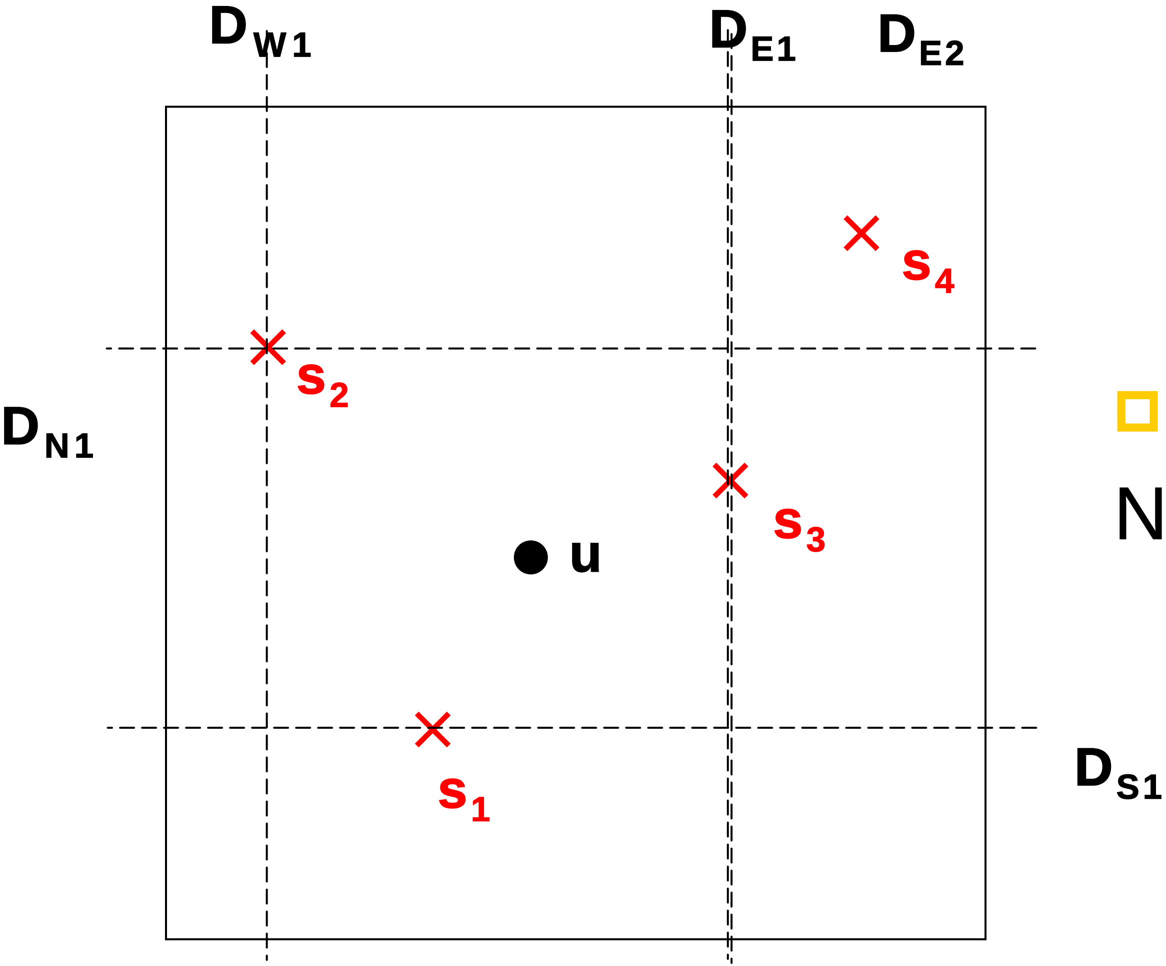
$$GGC(M) = \sum_{s \in S} Area(MBR(\{s\} \cup M(s)))$$

Related work: Query Privacy in LBS

- Relies on Spatial K-anonymity as well
  - Gedik & Liu, ICDCS ’05
  - Mokbel et al., VLDB ’06
  - Kalnis et al., TKDE ’07
  - ...
- But anonymizes a **single** query
  - Equivalent to |S|=1
- In our problem |S|>>1
  - More difficult to solve

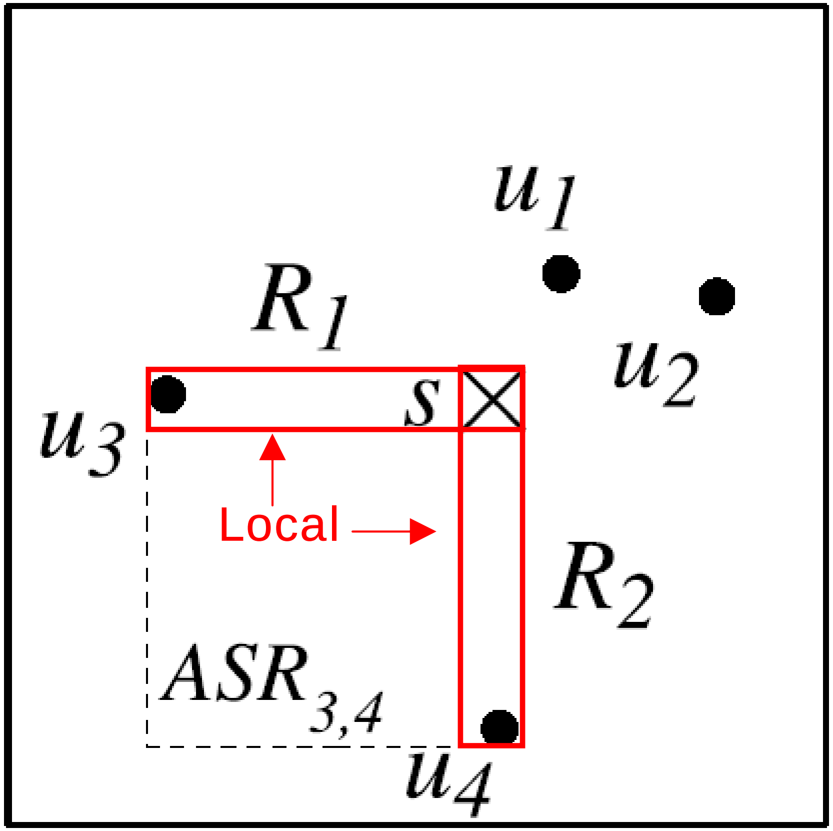
Related work: “Local” algorithm

[HXD+07] Hu H., Xu J., Du J., Ng J.K.Y., “Privacy Aware Location Publishing for Moving Clients”, TR, Hong Kong Baptist Univ., 2007 www.comp.hkbu.edu.hk/~haibo/privacy_join.pdf



- Benefit: NumOfSites/Area

Drawback of Local's Publication Format



2-by-2 publishing

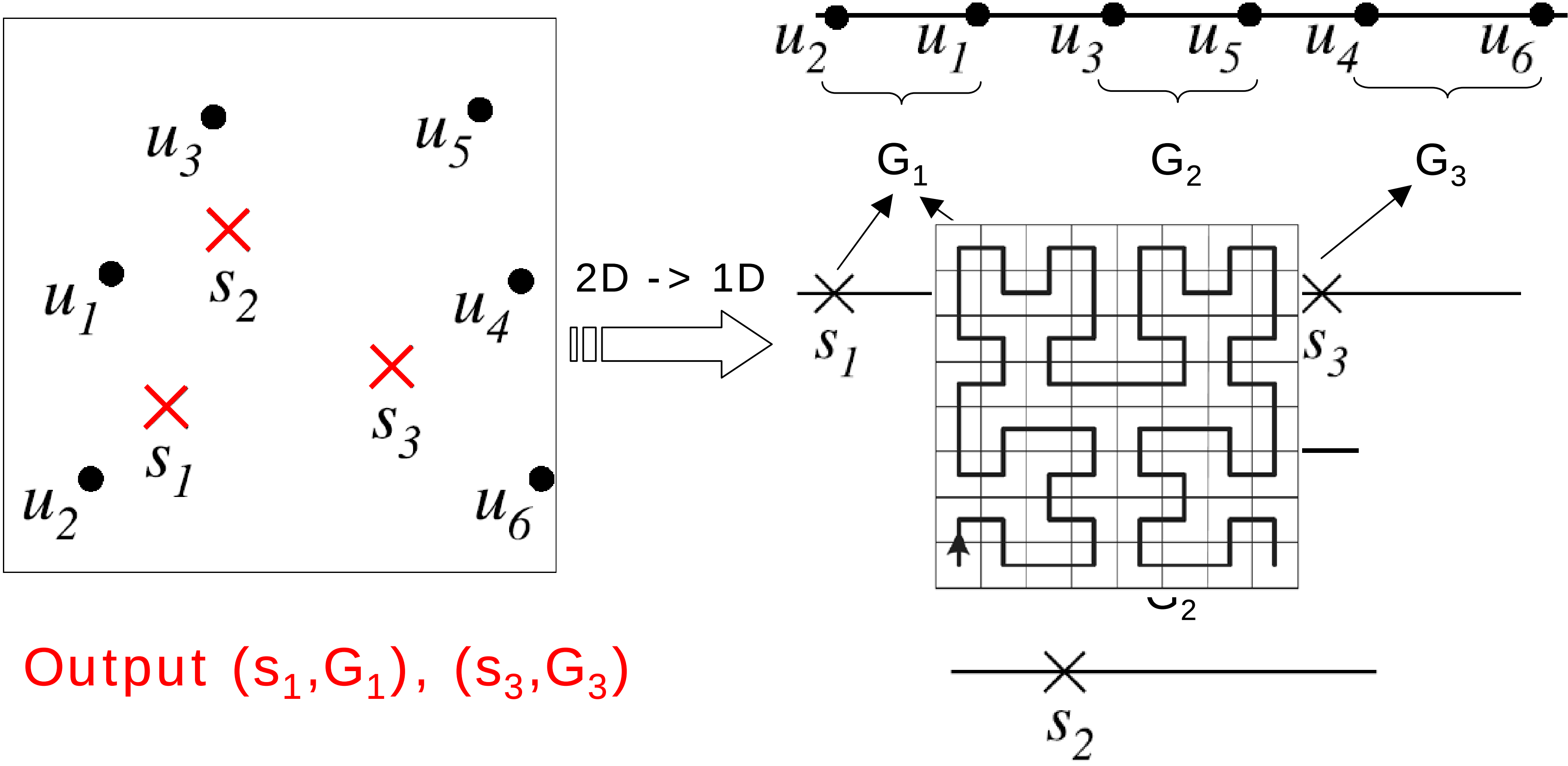
Our Approach
MK: Monochromatic *K*-anonymity

- Phase 1:
  - Transform user locations to 1D
  - Anonymization w.r.t. user set *U* ONLY
  - Linear algorithm*, 1-D optimal
  - User groups independent of sites *S*

- Phase 2:
  - Assign each anonymous group to nearest site
  - Resolve potential conflicts
    - E.g., choose assignment with minimum enlargement
  - Repeat until all sites are covered

[GKKM07] Ghinita G., Karras P., Kalnis P., Mamoulis N. “Fast Data Anonymization with Low Information Loss”. In Proc. of VLDB 2007

MK Example



Output (s_1,G_1), (s_3,G_3)

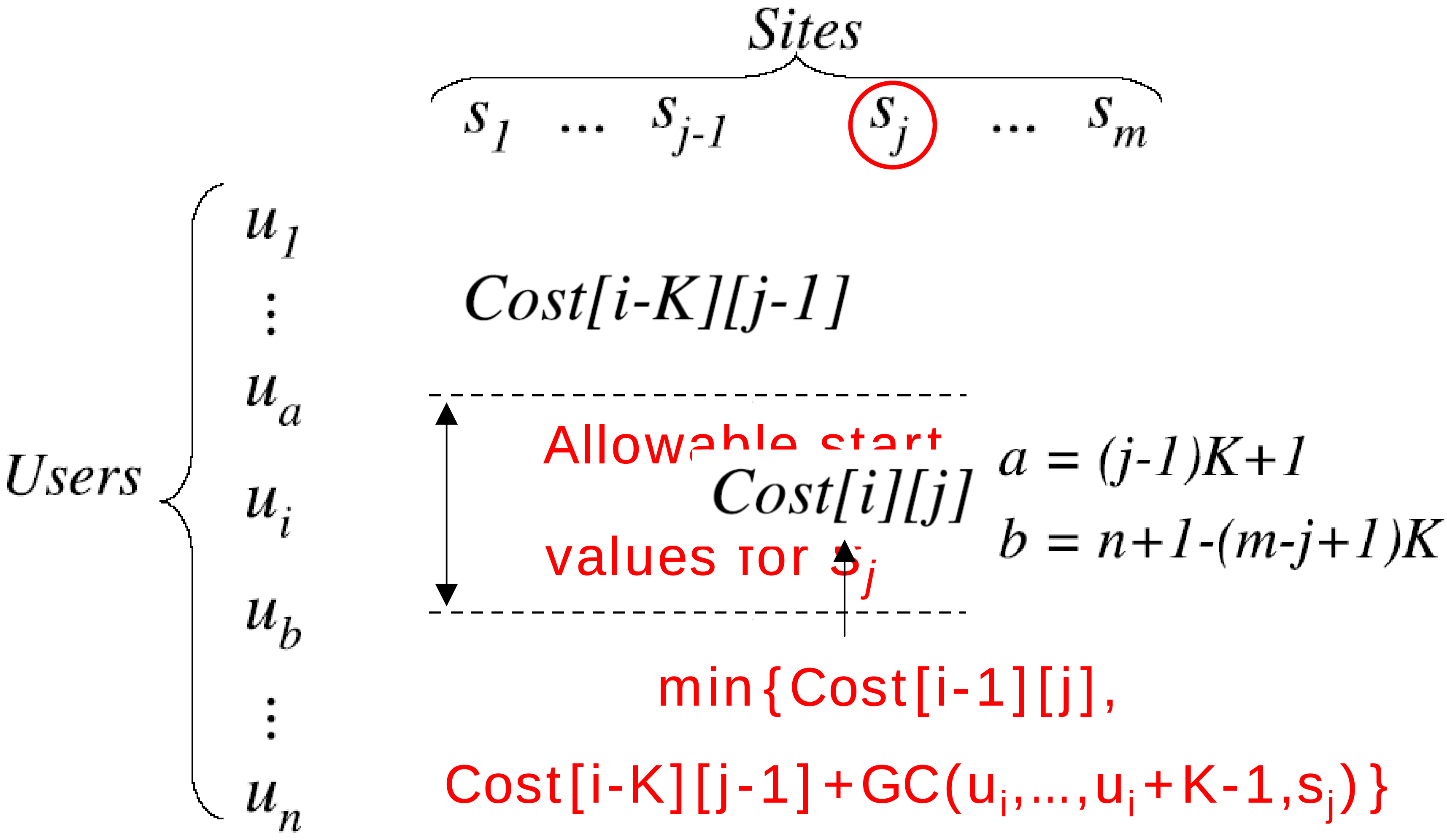
Output (s_2,G_2)

Our Approach
BK: Bichromatic *K*-anonymity

- Properties of optimal mapping in 1-D
  1. Each ASR/group has exactly *K* users
  2. ASRs have consecutive users in 1-D order
  3. Groups do not overlap in 1-D order

- Bichromatic clustering of U and S
  - Each cluster has 1 site and *K* users
  - DP algorithm, linear in *K* and |*U*|

BK Example



Allowable start values for s_j

$a = (j-1)K+1$

$b = n+1-(m-j+1)K$

min{Cost[i-1][j],

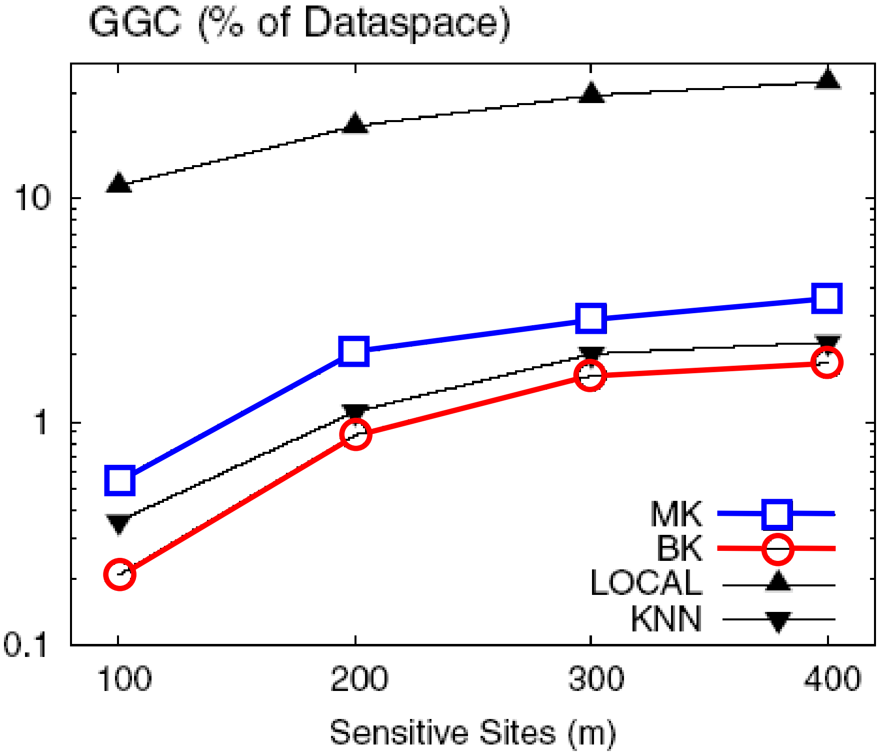
Cost[i-K][j-1]+GC(u_i,…,u_i+K-1,s_j)}

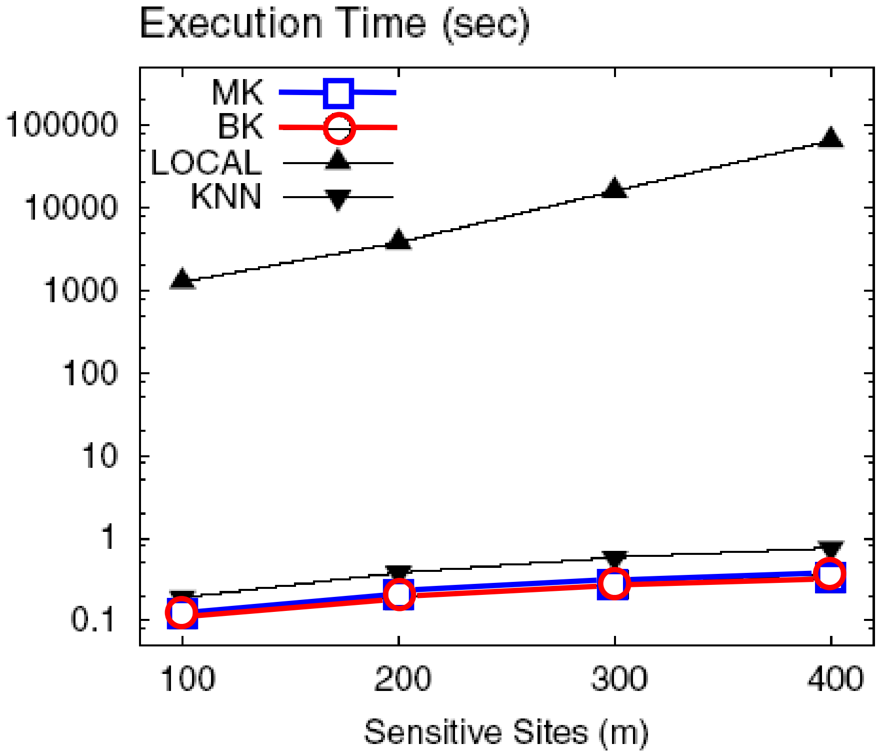
Experimental Evaluation

- Naïve competitor: **K-Nearest-Neighbors**
- NA Dataset (569120 locations)
  - *U* and *S* randomly sampled from NA
- Performance metrics:
  - Anonymization overhead (CPU time)
  - Generalization Cost

$$GGC(\mathcal{M}) = 100 \cdot \frac{\sum_{s \in S} Area(MBR(\mathcal{M}(s) \cup \{s\}))}{DomainArea}\%$$

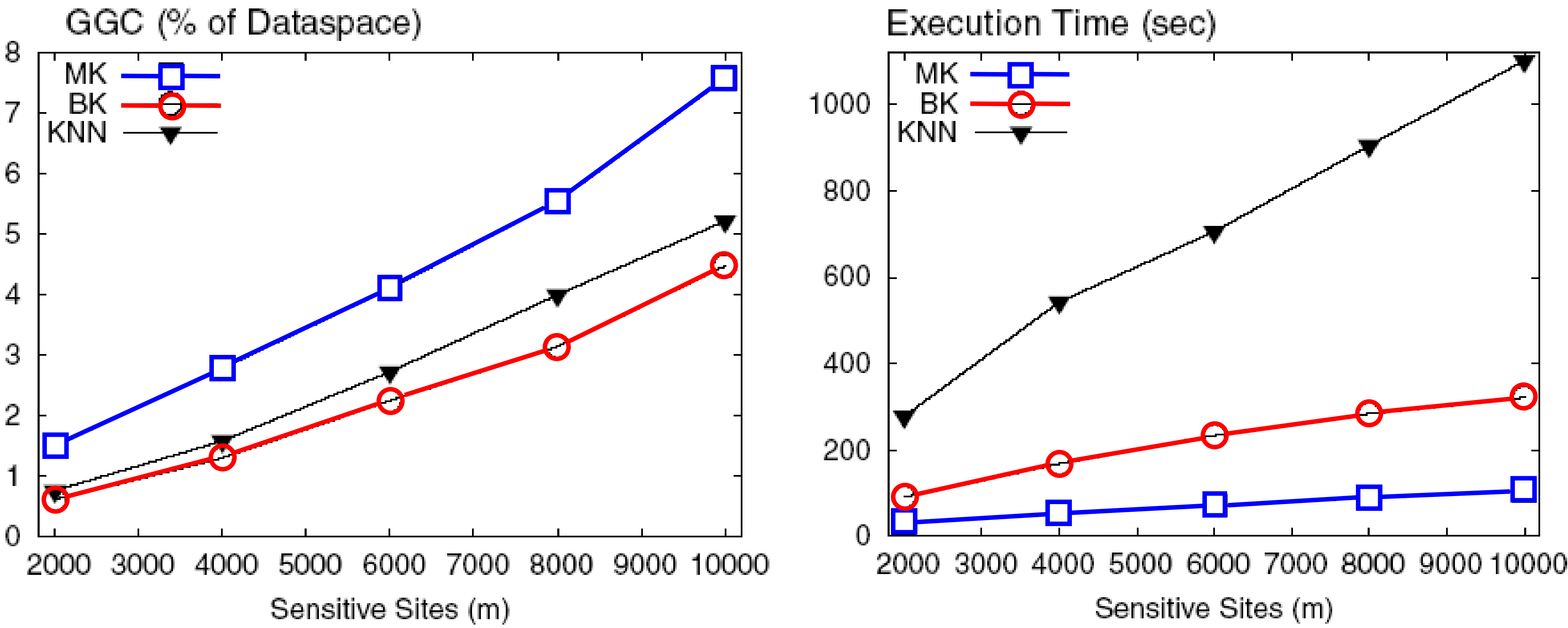
Comparison with “Local”





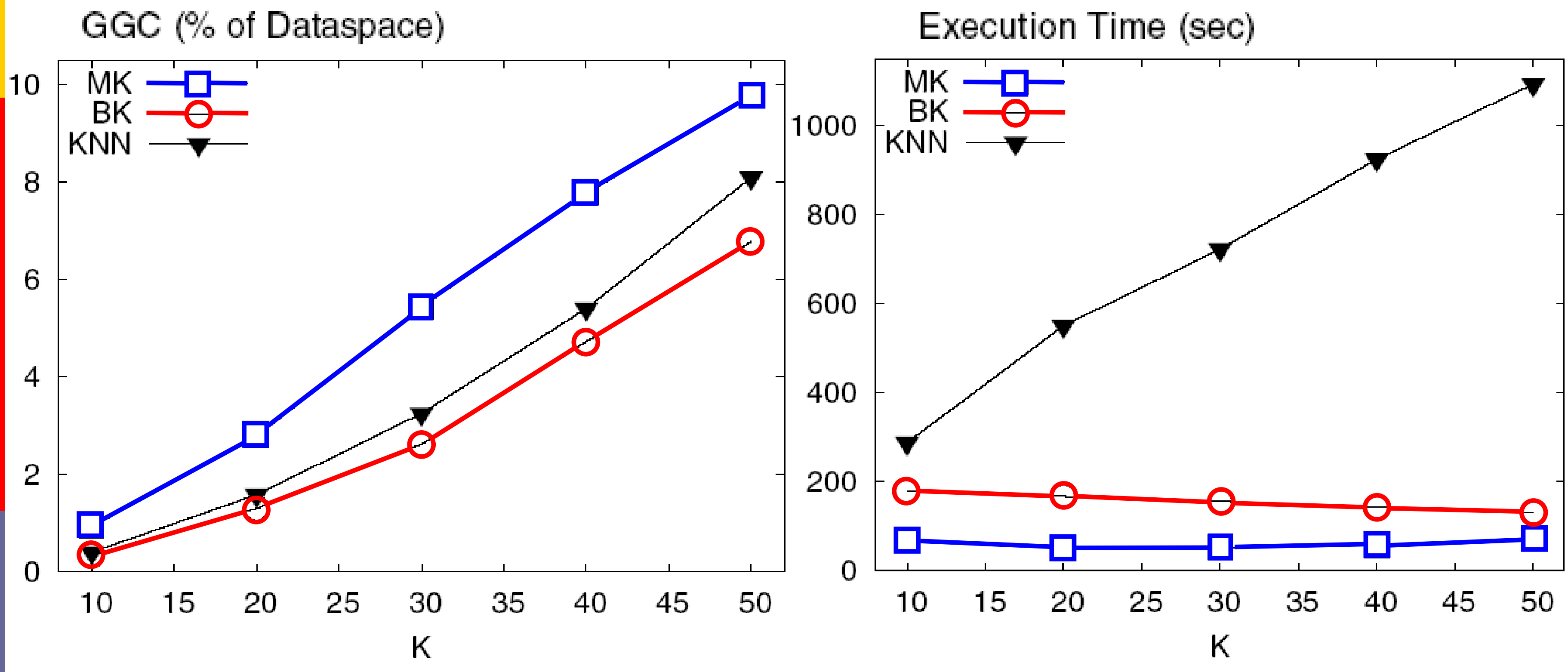
N=10,000 Users

Variable number of sites



N=569,120 Users

Variable *K*



N=569,120 Users

Points to Remember

- Publication of user locations in the proximity of sensitive sites
  - More difficult than Query-privacy in LBS
- “Local” algorithm
  - Very Slow
  - Bad quality, if a secure publishing format is used
- Naïve KNN
  - Also slow
- Our algorithms: MK, BK
  - Fast & Accurate

Bibliography on LBS Privacy

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