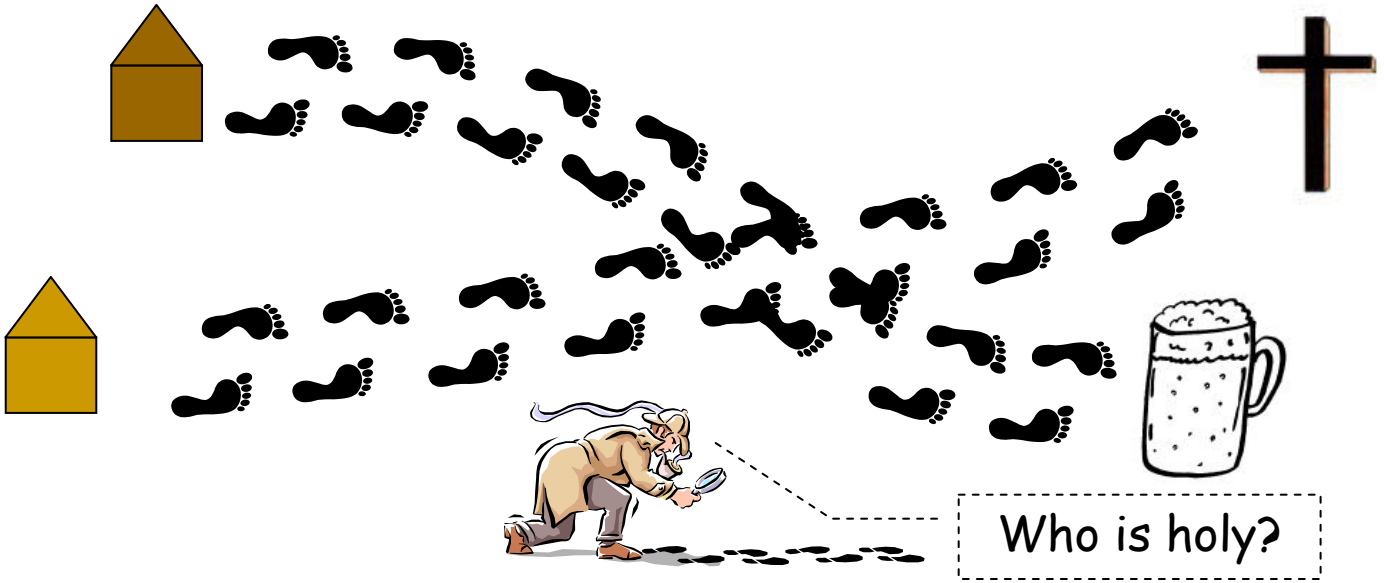
Protecting Location Privacy Through Path Confusion

Baik Hoh (baikhoh@winlab)

Marco Gruteser (gruteser@winlab)

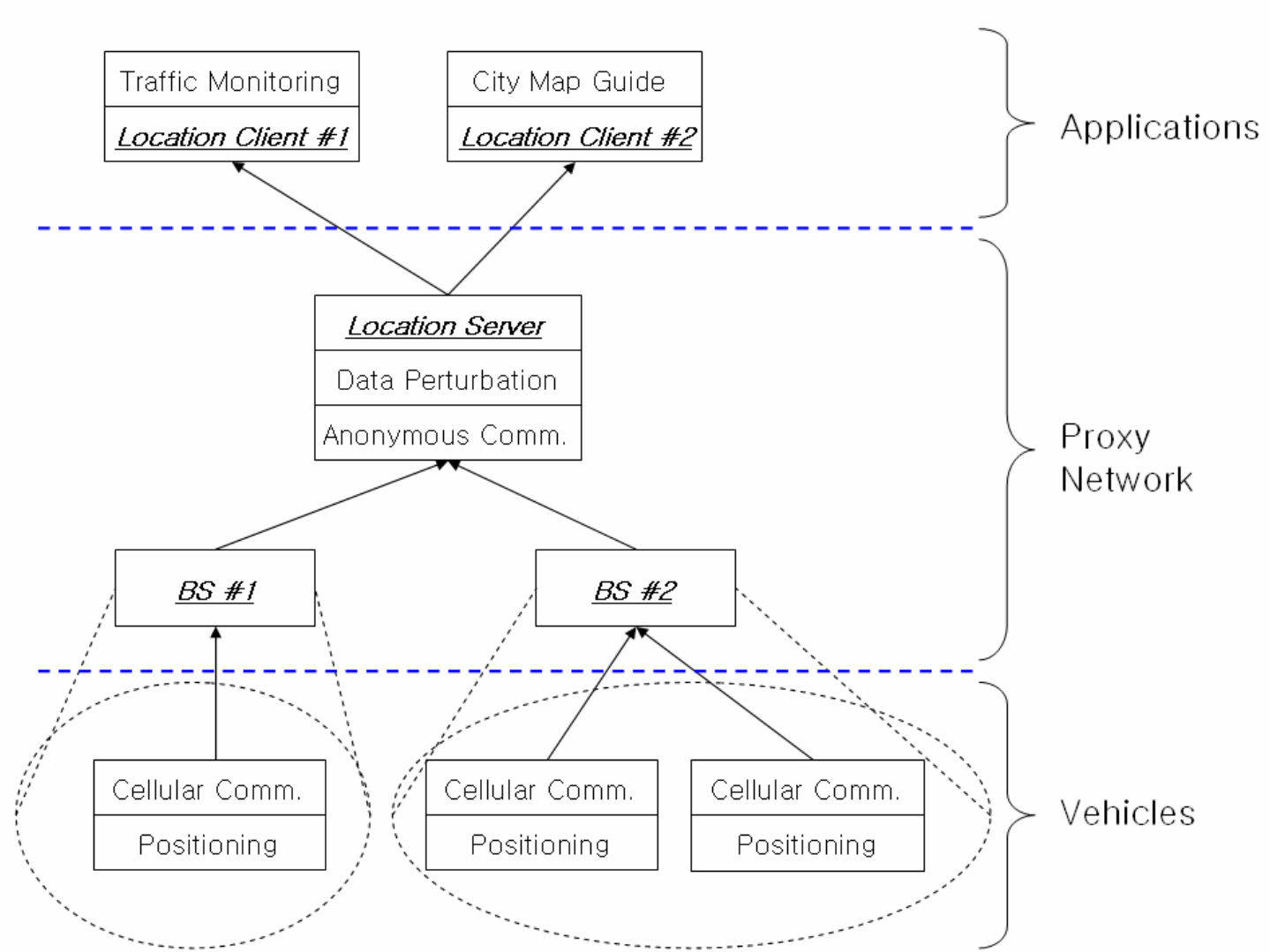
“Big brother is seeing you…”



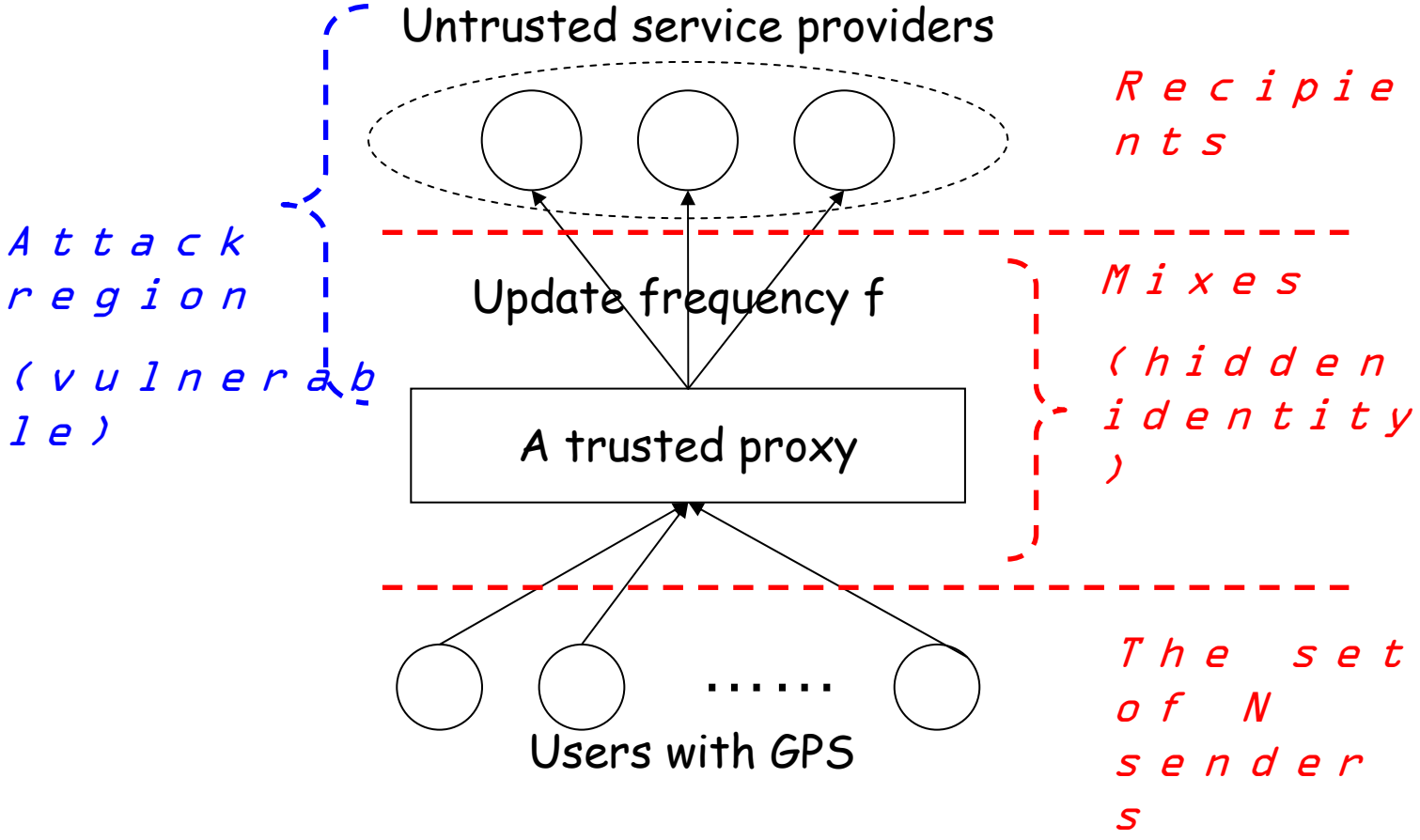
The term "Big Brother" refers to the government of Oceania, the setting of George Orwell's classic "1984."

- IBM signs $125 mil. deal to build nationwide automobile monitoring system in UAE, the largest telematics deal in history. (*InformationWeek, April 15, 2005*)

System Model

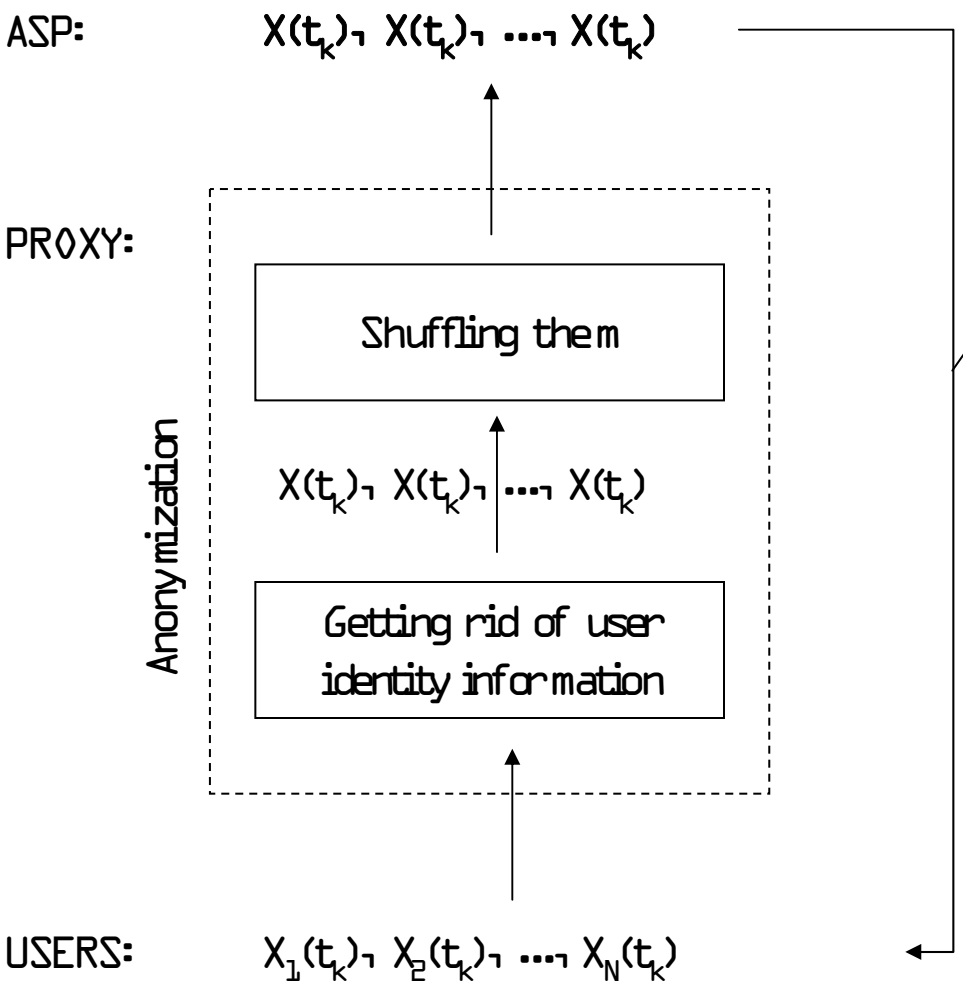


Location Privacy Framework



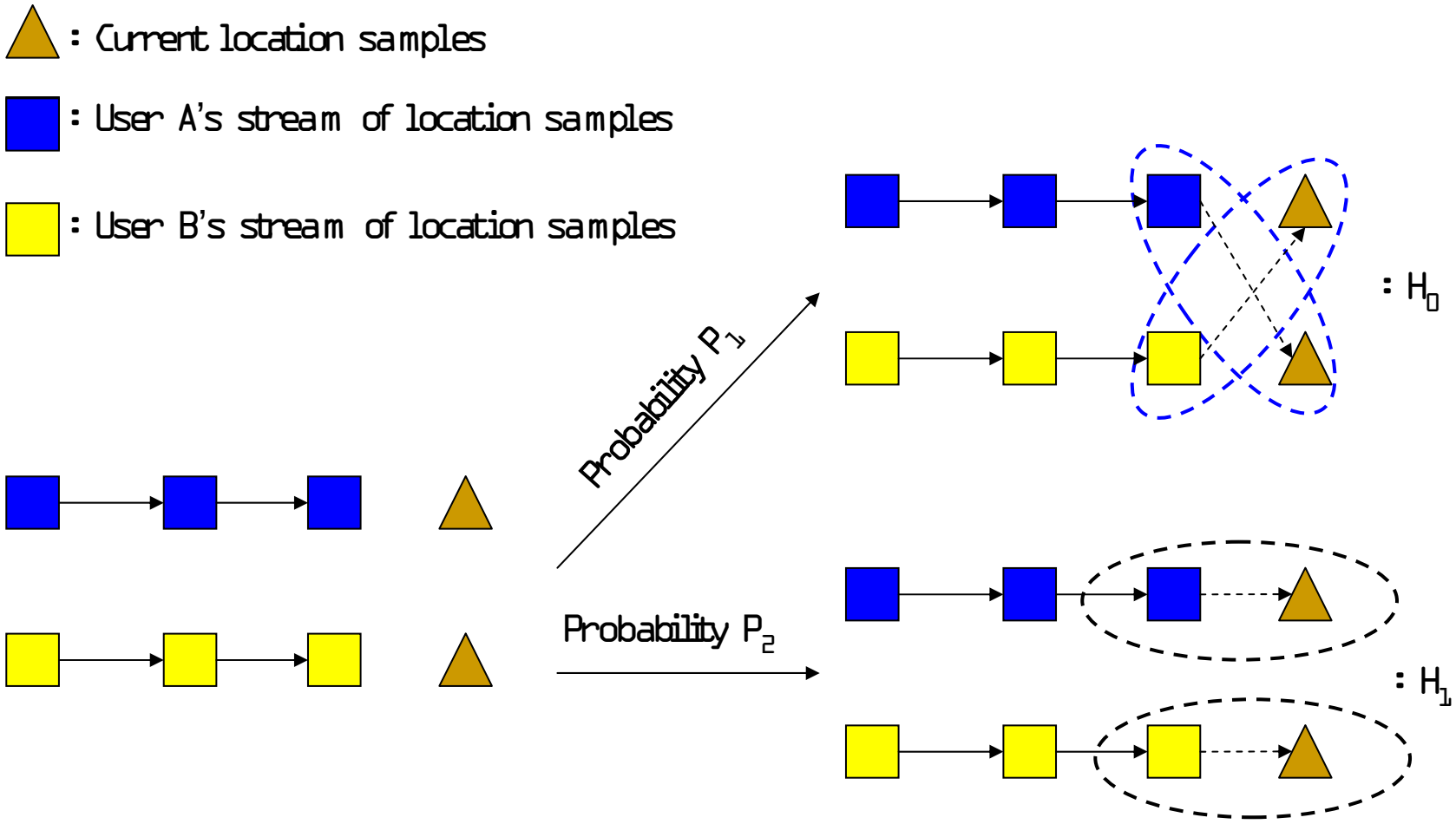
- The role of a trusted proxy: anonymization gets rid of user identity and shuffles location packets

Attack Model



Adversary tries to assign each location to each user → Well-studied algorithm, 'data association problem' → Reid's multi-hypothesis tracking algorithm

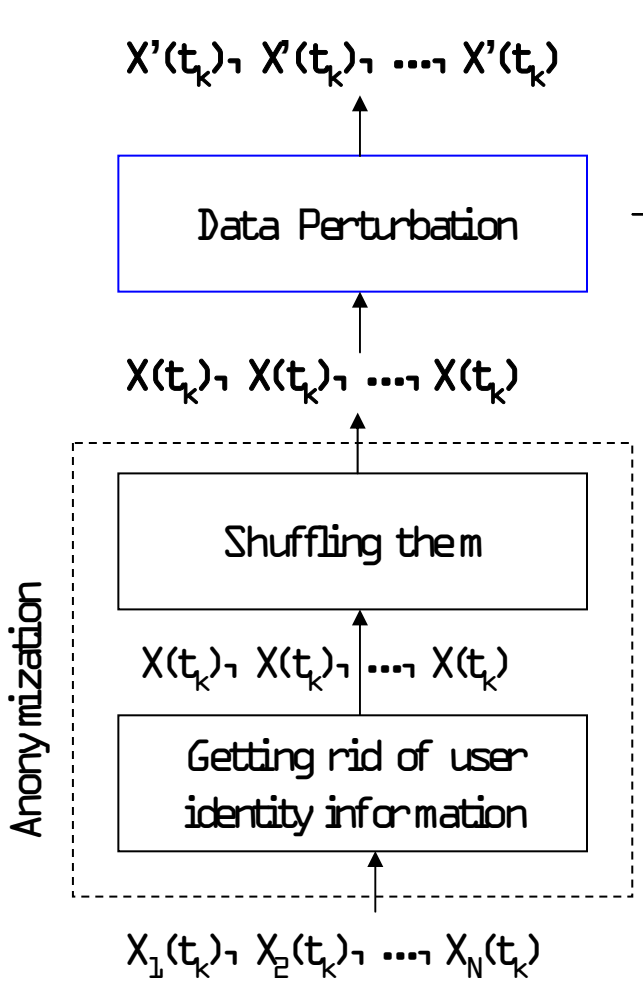
Attack Model (Example)



Approach

- Research statement
  - Naive anonymization is not sufficient for low user densities
  - How can we maximize location privacy while using anonymous location-based applications?
- How to preserve privacy against an adversary?
  - Adding spatial uncertainty: path perturbation
  - Formulating the problem as a constrained optimization problem

Approach (cont.)



We find the optimal set of values {X'(tk), X'(tk), ..., X'(tk)} which maximizes *location privacy* at the bearable *loss of quality of service*. (tradeoff)

Metrics and Optimization formulation

- We can formulate perturbation as a constrained nonlinear optimization problem.

$$\max_{\forall \widetilde{x_n}(k), \forall \widetilde{y_n}(k)} \sum_{i=1}^{I} p_i(k) d_i(k)$$

such that for every user

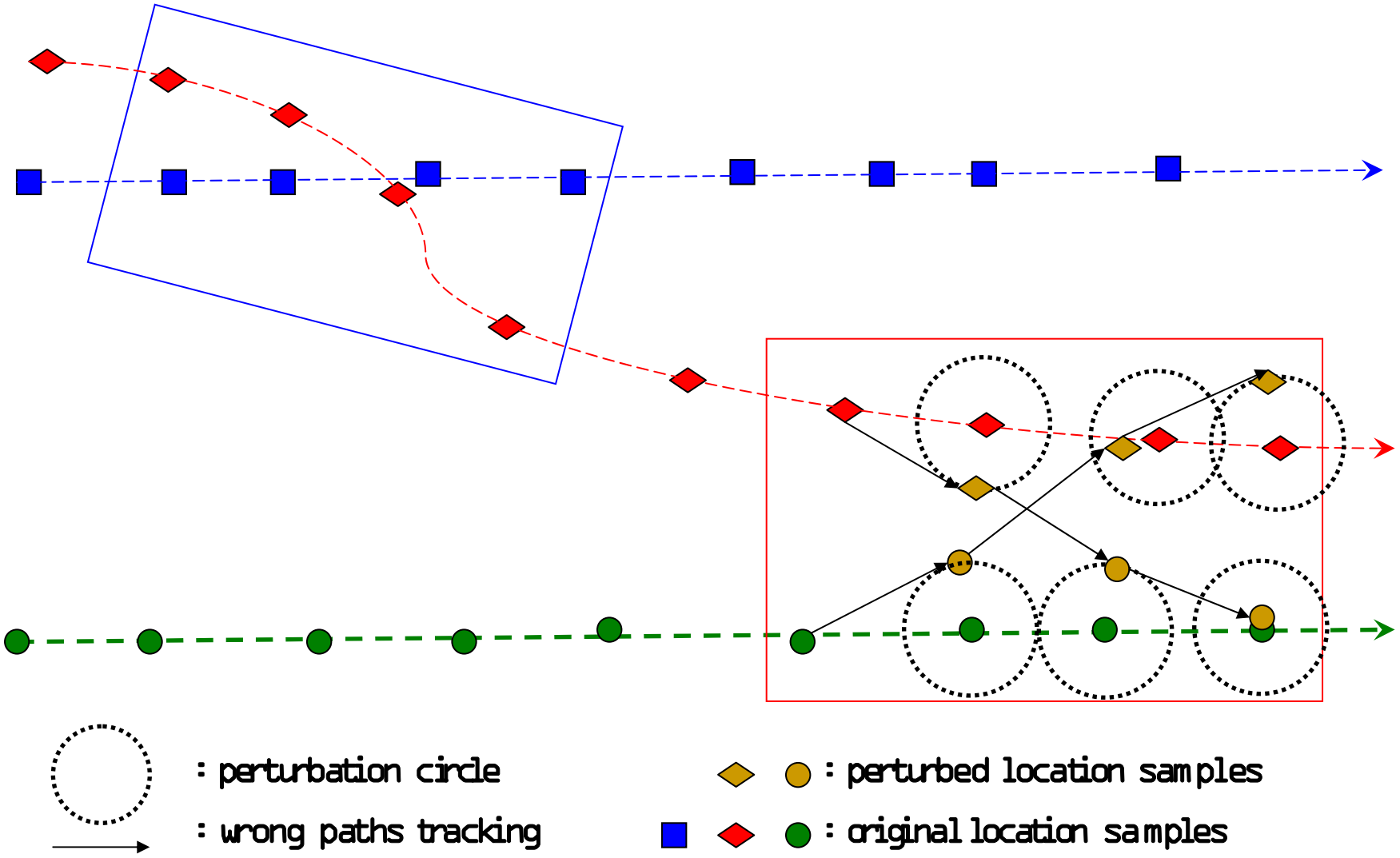
$$(\widetilde{x_n}(k) - x_n(k))^2 + (\widetilde{y_n}(k) - y_n(k))^2 \leq R^2$$

where $d_i(k)$ and $p_i(k)$ are described by the following equations.

$$d_i(k) = \sum_{n=1}^{N} \sqrt{(\widetilde{x_{m_i(n)}}(k) - x_n(k))^2 + (\widetilde{y_{m_i(n)}}(k) - y_n(k))^2}$$

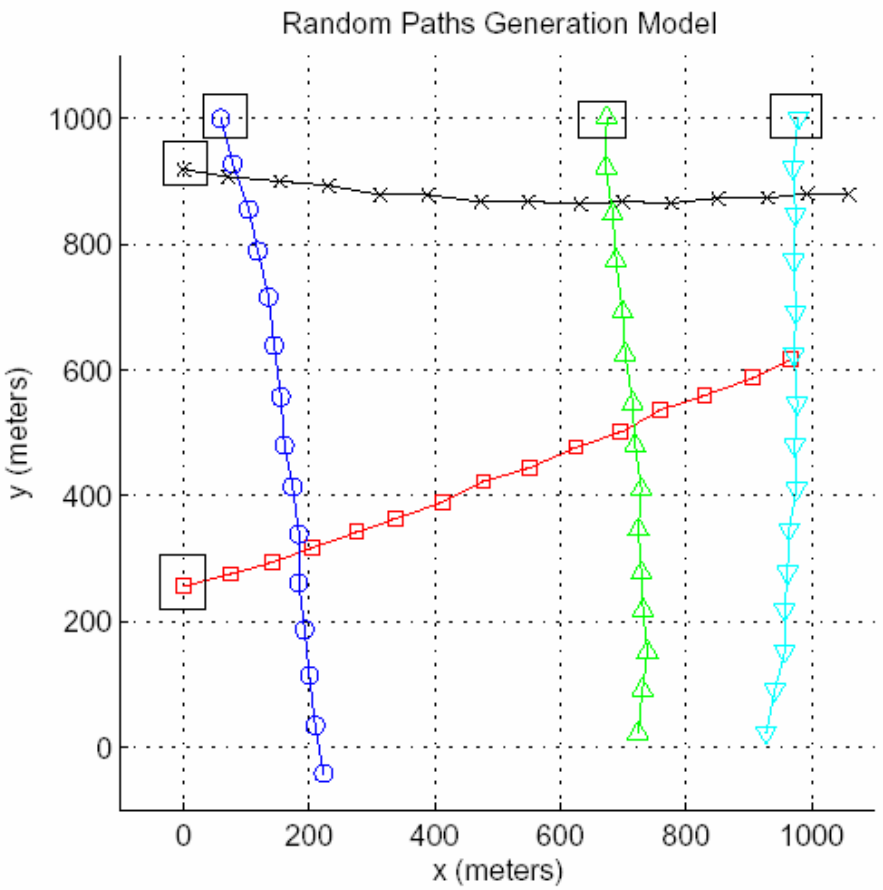
$$P_i^k \equiv P(\Omega_i^k \mid Z^k) \approx \prod_{n=1}^{N} f(\widetilde{x_{m_i(n)}}(k), \widetilde{y_{m_i(n)}}(k))$$

Path Perturbation/Segmentation



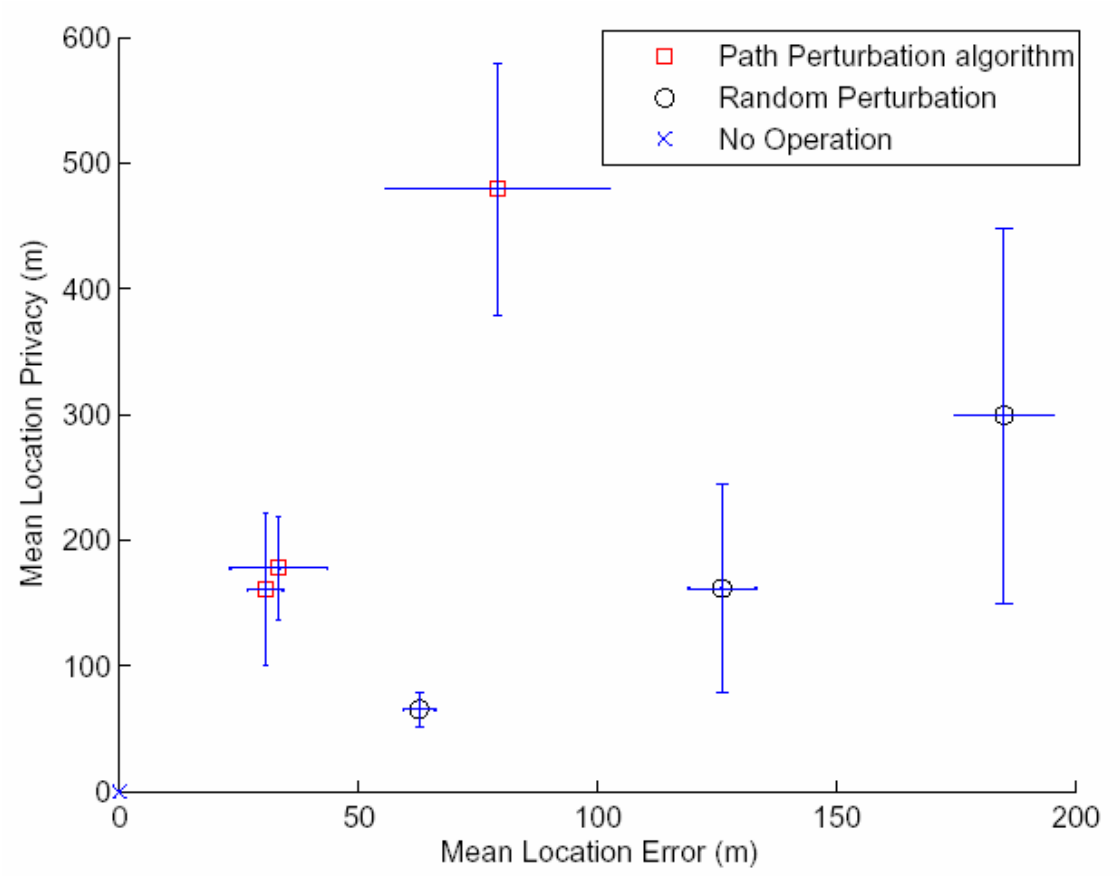
Experiments setup

- User density: 5/km^2 (=12.8/mile^2)
- Reporting frequency: T_f=5 sec
- Random Movement Model
  - Process model
  - Observation model



Performance evaluation

- Baseline to compare: random perturbation



Discussion & Conclusion

- Path Perturbation algorithm:
  - can limit the tracking duration
  - can improve location privacy with a lower mean location error
- Results dependent on
  - User density
  - Tracking algorithm
  - Characteristics of the original traces
- Future work:
  - Reduce the computation overhead (scalability)
  - Validate the results on datasets collected from real-life applications (feasibility)

Questions and Comments

Thank you