A Predictive Differentially-Private Mechanism for Mobility Traces

Marco Stronati
marco@stronati.org
joint work with
K. Chatzikokolakis and C. Palamidessi

Location Based Service

Scope

$$x \quad \longrightarrow \quad \mathcal{M} \quad \longrightarrow \quad z$$

Scope

$$x \longrightarrow \mathcal{M} \longrightarrow z$$

Privacy

through reduced accuracy

Utility

accuracy of reported location

Scope

$$x \longrightarrow \mathcal{M} \longrightarrow z$$

Privacy

through reduced accuracy

Utility

accuracy of reported location

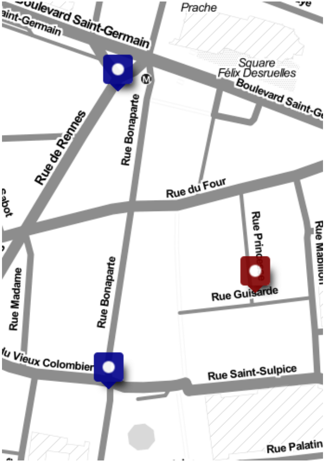
Contribution

in traces with considerable correlation we provide better utility

Privacy Definition

Geo-indistinguishability

$$d_{\mathcal{P}}(M(x), M(x')) \leq \epsilon \cdot d(x, x') \qquad \forall x, x'$$

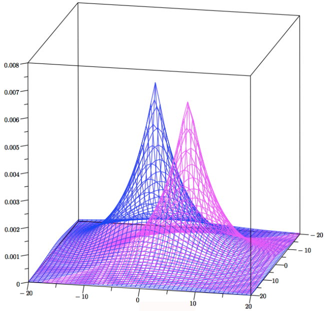


Andrés, Bordenabe, Chatzikokolakis, Palamidessi: Geo-indistinguishability: differential privacy for location-based systems. In: Proc. of CCS, ACM (2013) 901–914

Privacy Mechanism

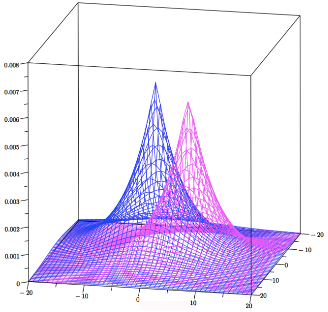
Noise mechanism

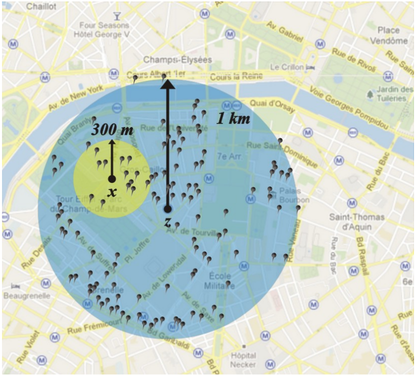
$$N(\epsilon_N)$$



Noise mechanism

$$N(\epsilon_N)$$





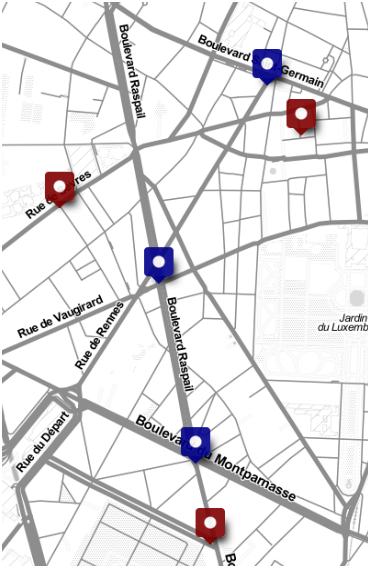
Mobility Traces

Independent Mechanism

$IM(\bar{x})$ that uses $N(\epsilon_N)(x)$ is

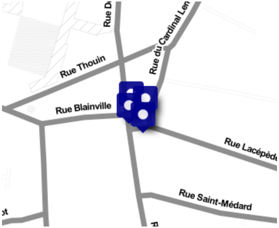
$$n \cdot \epsilon_N \ d\text{-private}$$

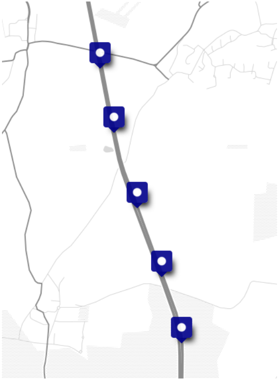
- works on *any* trace (including random teleporting)
- budget is linear with the length of the trace



Correlation

- real traces are strongly correlated
- not every point has the same value





Predictive Mechanism (broken)

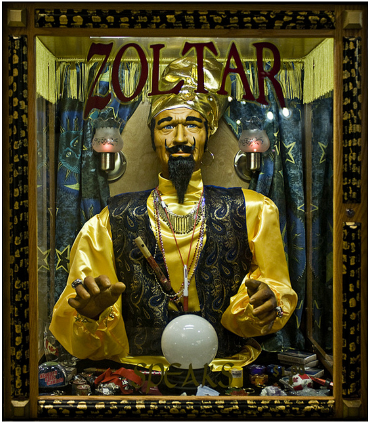
Predictive Mechanism (broken)

Equip the noise mechanism with

- a *prediction function*
- a test function with a threshold l

Cost

- *easy* points are free
- *hard* points cost ϵ_N



Predictive Mechanism (broken)

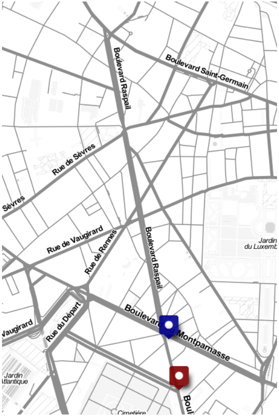
Predictive Mechanism (broken)

Equip the noise mechanism with

- a *prediction function*
- a test function with a threshold l

Cost

- *easy* points are free
- *hard* points cost ϵ_N



Predictive Mechanism (broken)

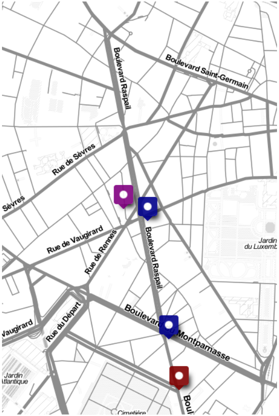
Predictive Mechanism (broken)

Equip the noise mechanism with

- a *prediction function*
- a test function with a threshold l

Cost

- *easy* points are free
- *hard* points cost ϵ_N



Predictive Mechanism (broken)

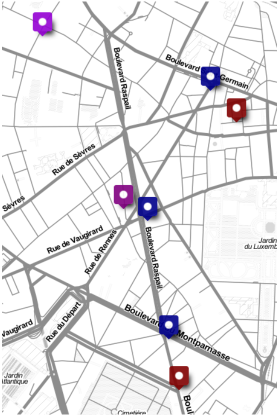
Predictive Mechanism (broken)

Equip the noise mechanism with

- a *prediction function*
- a test function with a threshold l

Cost

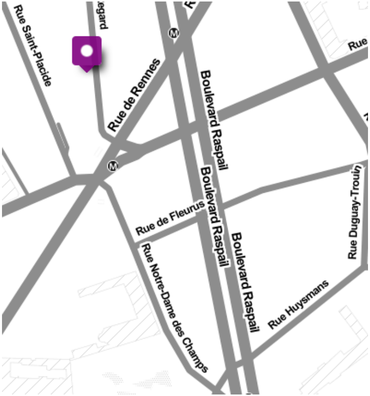
- *easy* points are free
- *hard* points cost ϵ_N



Testing for accuracy

Deterministic test

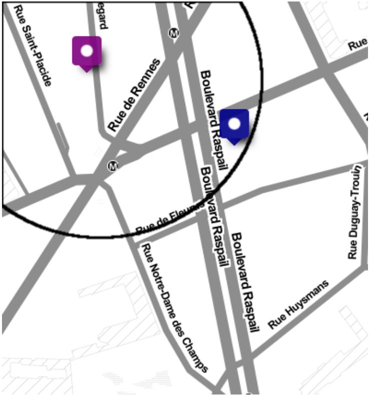
breaks d-privacy: two close secrets always report different observables



Testing for accuracy

Deterministic test

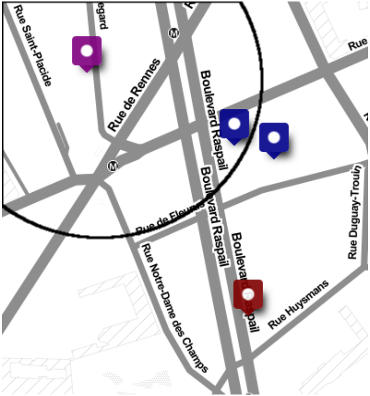
breaks d-privacy: two close secrets always report different observables



Testing for accuracy

Deterministic test

breaks d-privacy: two close secrets always report different observables



Testing for accuracy

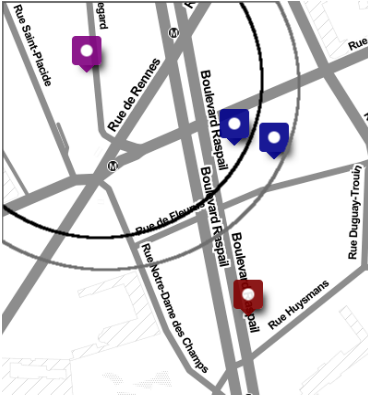
Deterministic test

breaks d-privacy: two close secrets always report different observables

D-Private test

$$\Theta(\epsilon_\theta, l)$$

adds again laplacian noise on the distance between secret and prediction



Testing for accuracy

Deterministic test

breaks d-privacy: two close secrets always report different observables

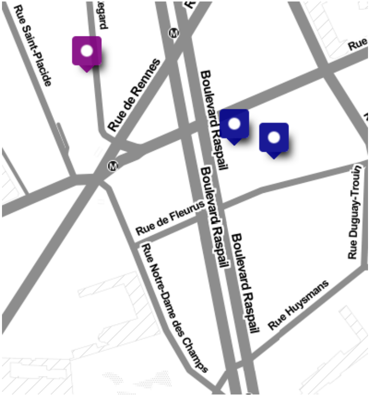
D-Private test

$$\Theta(\epsilon_\theta, l)$$

adds again laplacian noise on the distance between secret and prediction

Skip the test

testing is still linear in n



Predictive Mechanism

Predictive Mechanism

$PM(\epsilon_\theta, \epsilon_N, l)$

- prediction function
- d-private test $\Theta(\epsilon_\theta, l)$
- noise mechanism $N(\epsilon_N)$

Results

- the mechanism is indeed d-private
- the budget used at each step is ϵ_θ (easy) or $\epsilon_\theta + \epsilon_N$ (hard)
- global budget depends on the run (on the trace)

Budget Managers

Parameters

- Local: $(\epsilon_\theta, \epsilon_N, l)$
- Global: (ϵ, α, n)
- Budget Manager: Global $\rightarrow$ Local

Budget Managers

Parameters

- Local: $(\epsilon_\theta, \epsilon_N, l)$
- Global: (ϵ, α, n)
- Budget Manager: Global $\rightarrow$ Local

Privacy

fixed ϵ we define two strategies

Fixed Accuracy

What is saved is spent to increase n

Fixed Rate

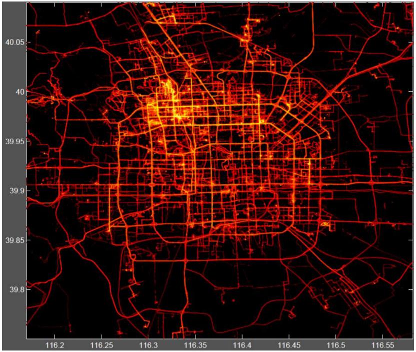
What is saved is spent to decrease α

Parrot prediction - simple yet effective

Parrot prediction - simple yet effective

repeats the last observable

Geolife and TDrive from Microsoft

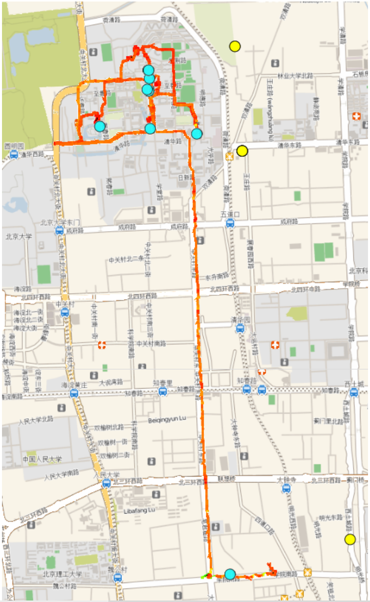


Sampling

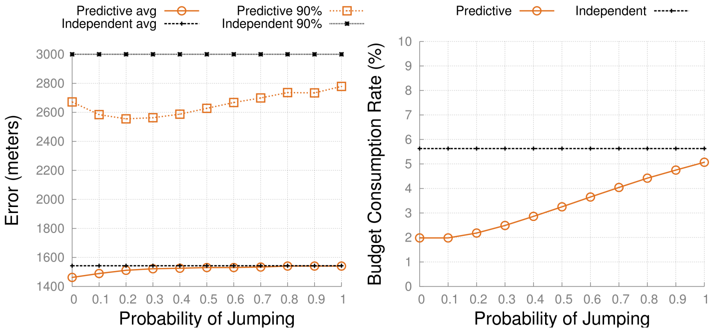
Sampled the traces with different frequencies

- 1 minutes
- 1 hour (a *jump*)

- Original trace
- Sampled trace
- Reported trace

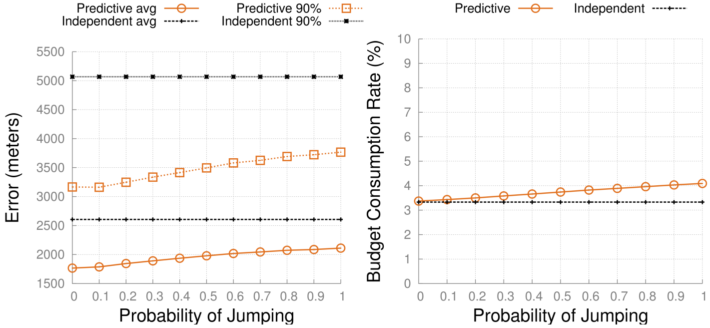


Experimental results



Geolife: Fixed Accuracy 3 km
with skip

Experimental results



Geolife: Fixed Rate 3.3%

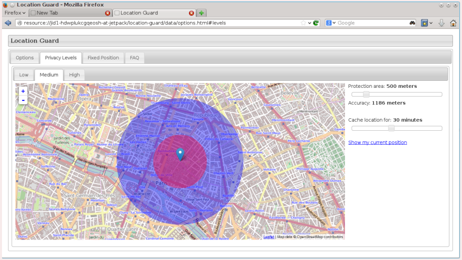
What to take home

- composition of private and deterministic components
- budget managers allows to move cost from privacy to accuracy or rate
- 99% predictive mechanism is reusable
- considerable correlation is needed to make up for the test cost

Thanks

Questions?

Location Guard for Chrome and Firefox



https://github.com/chatziko/location-guard