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

\[ x \rightarrow M \rightarrow z \]


Scope

\[
\begin{align*}
x & \rightarrow M & \rightarrow z
\end{align*}
\]

Privacy
through reduced accuracy

Utility
accuracy of reported location
Scope

Privacy
through reduced accuracy

Utility
accuracy of reported location

Contribution
in traces with considerable correlation we provide better utility
Privacy Definition

Geo-indistinguishability

\[ d_P(M(x), M(x')) \leq \epsilon \cdot d(x, x') \quad \forall x, x' \]

Privacy Mechanism

Noise mechanism

$N(\epsilon_N)$
Privacy Mechanism

Noise mechanism

\[ N(\epsilon_N) \]
Mobility Traces

Independent Mechanism

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

$n \cdot \epsilon_N$ d-private

- works on any trace (including random teleporting)
- budget is linear with the length of the trace
real traces are strongly correlated
not every point has the same value
Equip the noise mechanism with

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

Cost

- *easy* points are free
- *hard* points cost $\epsilon_N$
Equip the noise mechanism with
- a prediction function
- a test function with a threshold $l$

**Cost**
- *easy* points are free
- *hard* points cost $\epsilon_N$
Equip the noise mechanism with
- a prediction function
- a test function with a threshold $l$

Cost
- easy points are free
- hard points cost $\epsilon_N$
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 $\epsilon_N$
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

\[ \Theta(\varepsilon_\theta, l) \]
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

$PM(\epsilon_{\theta}, \epsilon_{N}, l)$
- 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 $\epsilon_{\theta}$ (easy) or $\epsilon_{\theta} + \epsilon_{N}$ (hard)
- global budget depends on the run (on the trace)
Parameters

- Local: \((\epsilon_\theta, \epsilon_N, l)\)
- Global: \((\epsilon, \alpha, n)\)
- Budget Manager: Global \(\rightarrow\) Local

Privacy fixed \(\epsilon\) we define two strategies

Fixed Accuracy
What is saved is spent to increase \(n\)

Fixed Rate
What is saved is spent to decrease \(\alpha\)
Budget Managers

Parameters
- Local: \((\epsilon_\theta, \epsilon_N, l)\)
- Global: \((\epsilon, \alpha, n)\)
- Budget Manager: Global \(\rightarrow\) Local

Privacy
fixed \(\epsilon\) we define two strategies

Fixed Accuracy
What is saved is spent to increase \(n\)

Fixed Rate
What is saved is spent to decrease \(\alpha\)
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
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
Questions?

Location Guard for Chrome and Firefox

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