Karlsruhe Institute of Technology

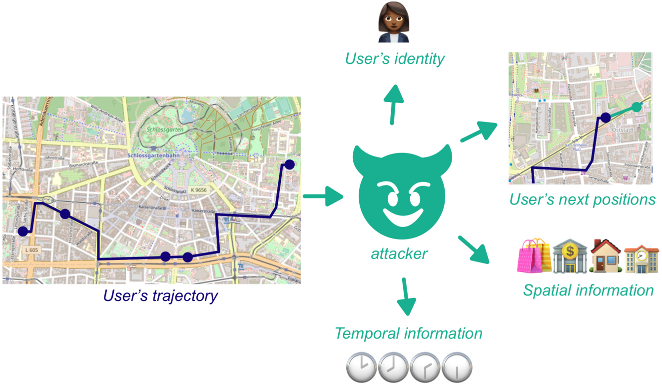
Publishing Trajectories: A Study on Privacy Protection

Patricia Guerra-Balboa, Àlex Miranda-Pascual,
T.Strufe, J.Parra-Arnau, J.Forné

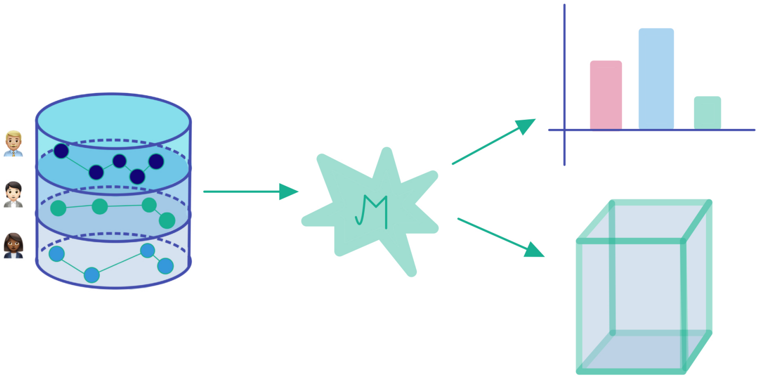
October 14, 2025

Motivation

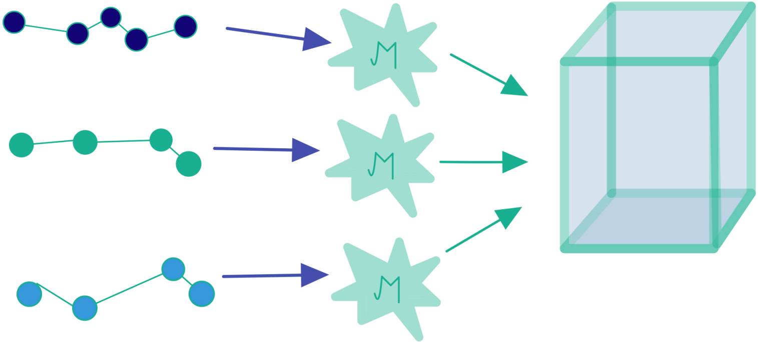
"Privacy is one the biggest problems in this new electronic age"- Andy Grove (former INTEL Ceo)



Statistical Disclosure Control

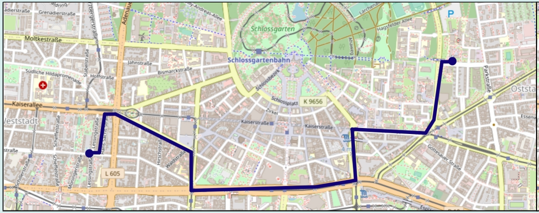


Statistical Disclosure Control

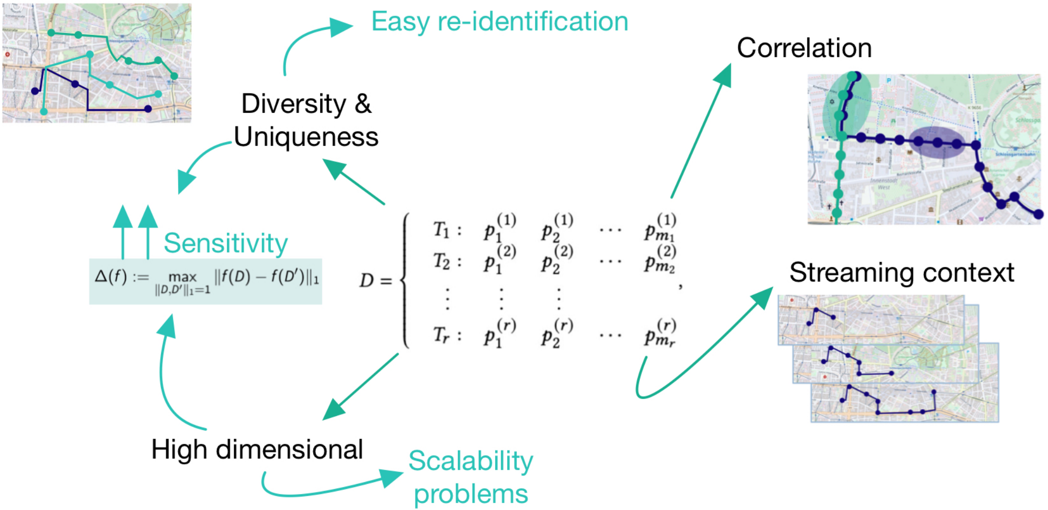


Model

Trajectorias

Raw Trajectories	Semantic Trajectories
	
$T = (x_1, y_1, t_1) \rightarrow \cdots \rightarrow (x_n, y_n, t_n)$	POI

Properties

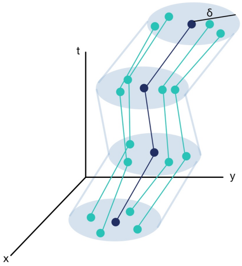


Easy re-identification
Diversity &
Uniqueness
Correlation
Sensitivity
$\Delta(f) := \max_{\|D,D'\|_1=1} \|f(D) - f(D')\|_1$
$D = \begin{cases} T_1: & p_1^{(1)} & p_2^{(1)} & \cdots & p_{m_1}^{(1)} \\ T_2: & p_1^{(2)} & p_2^{(2)} & \cdots & p_{m_2}^{(2)} \\ \vdots & \vdots & \vdots & & \\ T_r: & p_1^{(r)} & p_2^{(r)} & \cdots & p_{m_r}^{(r)} \end{cases},$
Streaming context
High dimensional
Scalability
problems

Privacy Notions in SDC

Syntactic Notions

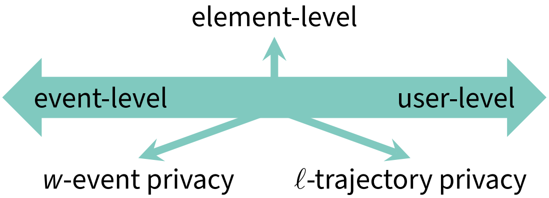
Database properties

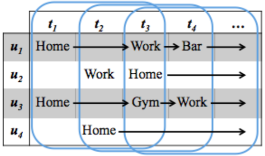


- *k*-anonymity
- *l*-diversity
- *t*-closeness
- Attribute Privacy

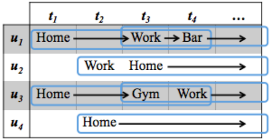
Semantic Notions

ϵ-differential privacy



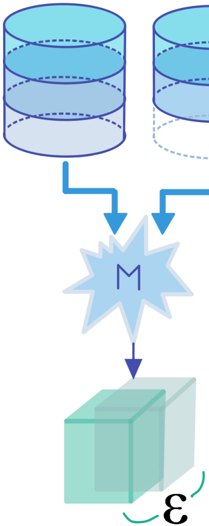


	t_1	t_2	t_3	t_4	...
u_1	Home		Work	Bar	
u_2		Work	Home		
u_3	Home		Gym	Work	
u_4		Home			



	t_1	t_2	t_3	t_4	...
u_1	Home		Work	Bar	
u_2		Work	Home		
u_3	Home		Gym	Work	
u_4		Home			

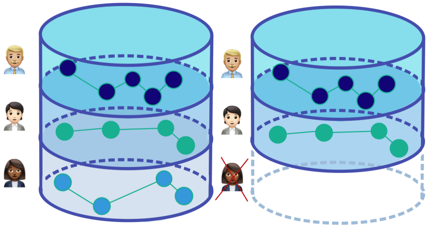
Differential Privacy



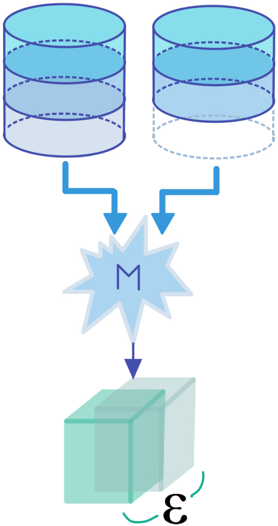
ε-Differential Privacy

A randomized algorithm M is said to be ε-*differentially private* if for all *neighboring* databases D, D' and all $\mathcal{S} \subseteq Range(M)$,

$$\mathbb{P}\{M(D) \in \mathcal{S}\} \leq e^{\varepsilon}\, \mathbb{P}\{M(D') \in \mathcal{S}\}.$$

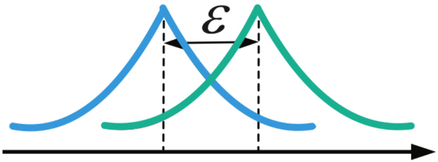


Differential Privacy



Privacy Loss (by observing r)

$$\mathcal{L}^{r}_{M(D)||M(D')} = ln\left(\frac{\mathbb{P}(M(D)=r)}{\mathbb{P}(M(D')=r)}\right)$$



Sensitivity

ℓ_1-sensitivity

The ℓ_1-sensitivity of a function $f : \mathbb{N}^{|\mathcal{X}|} \rightarrow \mathbb{R}^n$ is:

$$\Delta(f) := \max_{\|D,D'\|_1=1} \|f(D) - f(D')\|_1$$

UNBOUNDED SENSITIVITIES!!

outliers and huge noise

S-O-T-A analysis

Privacy Notion	Classification		Ref.	Correct DP notion	Mech.: Laplace	Mech.: Exponential	Properties: Considers time	Properties: Unb. loc. univ.	Properties: Realism	Utility Metrics: Total data preserv.	Close data preserv.: SM: *Euclidean*	Close data preserv.: SM: *Hausdorff*	Close data preserv.: SM: Other	Close data preserv.: Other	Statistics preserv.: *Loc. visit counts*	Statistics preserv.: *Freq. seq.*	Statistics preserv.: *Spatial density*	Statistics preserv.: Other	Realism assurance
ε-DP*	Noisy counts	Exploration tree	[107]		●	○	✗	✓	✗	●	○	○	○	●	●	●	●	●	●
ε-DP	Noisy counts	Exploration tree	[15, 16]		●	○	✗	✗	✗	○	○	○	○	○	●	●	○	○	○
			[30]		●	○	✗	✗	✗	○	○	○	○	●	●	○	○	○	○
		Sequence tree	[123]	✗	●	○	✗	✗	✗	○	○	○	○	○	○	○	○	●	○
			[121]	✗	●	○	✗	✗	✗	○	○	○	○	○	●	○	○	○	○
			[117]	✗	●	○	✗	✗	✗	○	○	○	○	○	○	○	○	●	○
		Trajectory count	[120]	✗	●	○	✗	✓	✓	○	○	○	○	○	○	●	○	○	○
	Noisy counts / Clustering	Tree + Markov Random centroid	[10]	✗	● ○	○ ○	✗	✓	✗	○	●	○	○	○	○	●	●	○	○
	Clustering	*k*-means	[19, 57]	✗	●	●	✗	✓	✗	○	○	●	○	○	○	○	○	○	○
			[70]	✗	●	●	✗	✓	✗	○	○	●	○	○	○	○	○	○	○
		Hilbert curves	[54]	✗	●	●	✗	✓	✗	○	○	●	○	○	○	○	○	●	○
		Universal clustering	[122]	✗	●	○	✗	✓	✗	○	●	○	○	○	○	○	○	○	○
$(0, \delta)$-DP	Sampling + interpolation		[91]		○	○	✗	✓	✗	○	●	○	●	○	○	○	○	○	○
ε-LDP	Perturbation		[24]		○	●	✓	✓	✓	○	○	○	●	●	●	○	○	○	○

Privacy Limitations

Inherent properties of trajectory data

Correlation
leads to
filtering attacks
&
physical models attacks
&
fake trajectory detection

Problems of current proposals

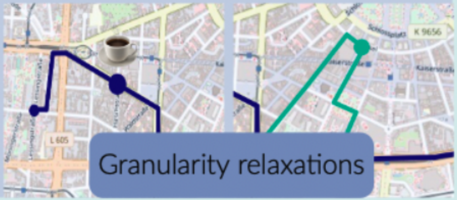
streaming context

Bayesian inference

not all dimensions are protected

Wrong proofs & sensitivities

semantics attribute disclosure



Granularity relaxations

Ulitility limitations

Inherent properties of trajectory data

Sparseness

leads to
Unavoidable data lost
&
high sensitivities

High dimensionality

Problems of current proposals

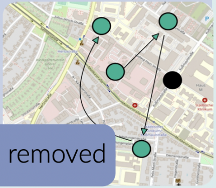
Impossible trajectories

Small universe of locations

Utility metrics are not representative

Ignoring the temporal dimension

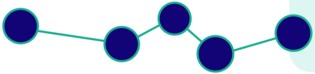
Weird trajectory patterns



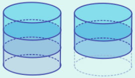
Stops are removed

Conclusions and Future Research

S.O.T.A
Trajectory anonymization

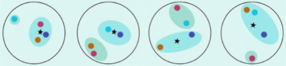


Data model + aspects to capture



Notions

Mechanisms



LIMITATIONS

Next steps

Mathematical understanding of correlation → New notions → New mechanisms