Karlsruhe Institute of Technology

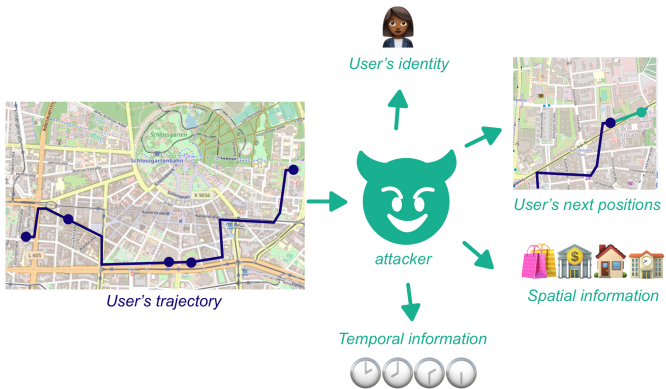
La Publicación de Trayectorias: un Estudio sobre la Protección de la Privacidad

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

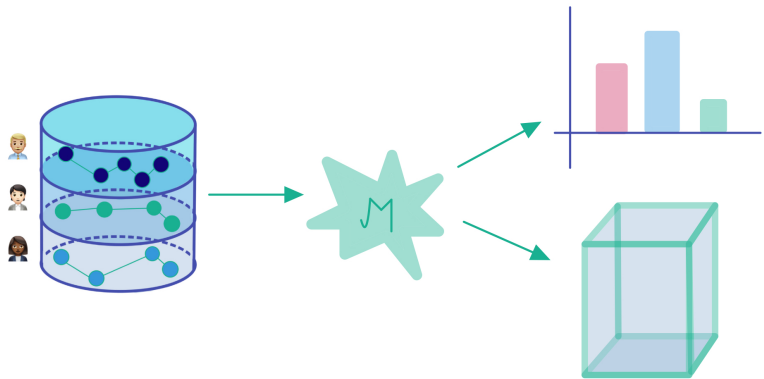
October 20, 2022

Motivation

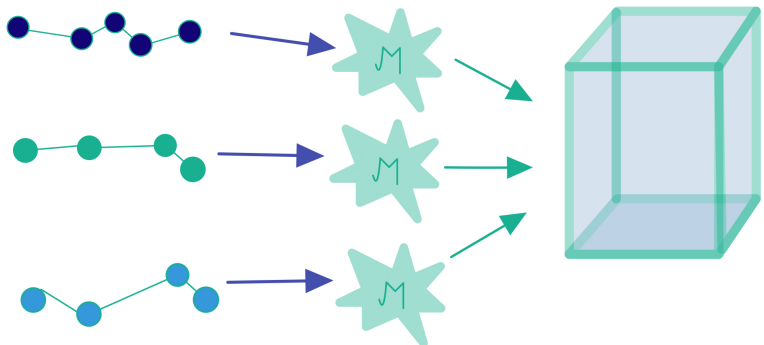
“Privacy is one the biggest problems in this new electronic age”- Andy Grove (former INTEL Ceo)



Statistical Disclosure Control



Statistical Disclosure Control



Model

Trajectorias

Raw Trajectories

$T = (x_1, y_1, t_1) \rightarrow \cdots \rightarrow (x_n, y_n, t_n)$

Semantic Trajectories

POI

Properties

Easy re-identification

Diversity & Uniqueness

Sensitivity

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

$$D = \begin{cases} T_1: & p_1^{(1)} & p_2^{(1)} & \cdots & p_{m_1}^{(1)} \\ T_2: & p_1^{(2)} & p_2^{(2)} & \cdots & p_{m_2}^{(2)} \\ \vdots & \vdots & \vdots & & \\ T_r: & p_1^{(r)} & p_2^{(r)} & \cdots & p_{m_r}^{(r)} \end{cases},$$

High dimensional

Scalability problems

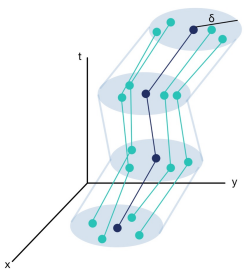
Correlation

Streaming context

Privacy Notions in SDC

Syntactic Notions

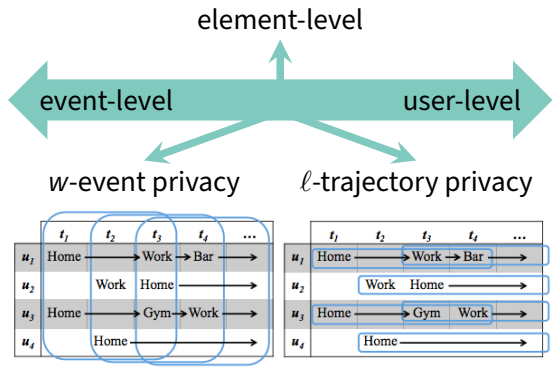
Database properties



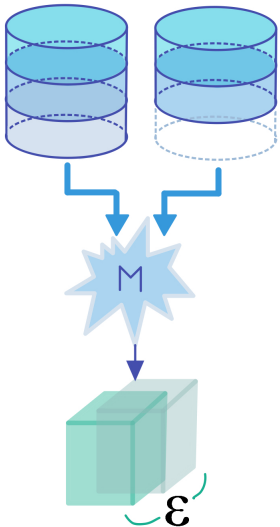
- *k*-anonymity
- *l*-diversity
- *t*-closeness
- Attribute Privacy

Semantic Notions

ϵ-differential privacy



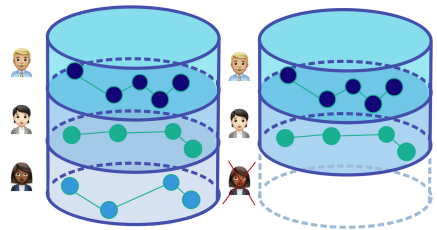
Differential Privacy



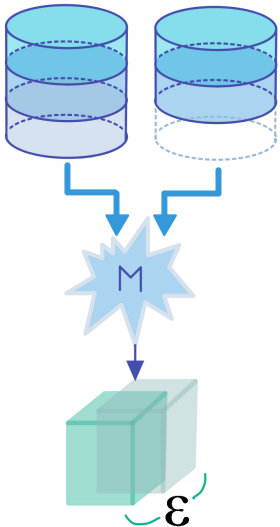
ε-Differential Privacy

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

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

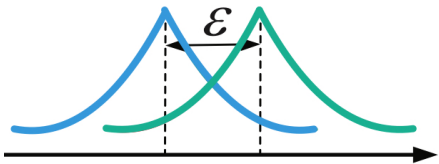


Differential Privacy



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 $f\colon \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^n$ 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	Laplace	Exponential	Considers time	Unb. loc. univ.	Realism	Total data preserv.	SM: *Euclidean*	SM: *Hausdorff*	SM: Other	Other	*Loc. visit counts*	*Freq. seq.*	*Spatial density*	Other	Realism assurance
					Mech.		Properties			Utility Metrics									
											Close data preserv.				Statistics preserv.				
ε-DP*			[107]		●	○	✗	✓	✗	●	○	○	○	●	●	●	●	●	●
ε-DP	Noisy counts	Exploration tree	[15, 16]		●	○	✗	✗	✗	○	○	○	○	○	●	●	○	○	○
			[30]		●	○	✗	✗	✗	○	○	○	○	●	●	○	○	○	○
		Sequence tree	[123]	✗	●	○	✗	✗	✗	○	○	○	○	○	○	○	○	●	○
			[121]	✗	●	○	✗	✗	✗	○	○	○	○	○	●	○	○	○	○
			[117]	✗	●	○	✗	✗	✗	○	○	○	○	○	○	○	○	●	○
		Trajectory count	[120]	✗	●	○	✗	✓	✓	○	○	○	○	○	○	●	○	○	○
		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

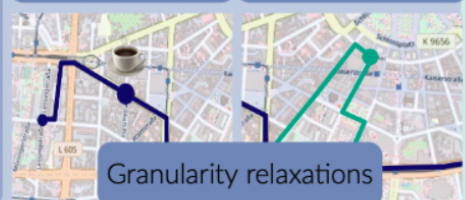
streaming context

Bayesian inference

not all dimensions are protected

semantics attribute disclosure

Wrong proofs & sensitivities



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