Anonymizing Trajectory Data: Limitations and Opportunities

Patricia Guerra-Balboa [1] Àlex Miranda Pascual [1,2] Javier Parra-Arnau [1,2] Jordi Forné [2] Thorsten Strufe [1]

[1]KASTEL Security Research Labs, Karlsruhe Institute of Technology [2]Dept. of Network Engineering, Universitat Politècnica de Catalunya

Motivation

The value of and interest into trajectory data are increasingly apparent. Traffic management, urban planning, and routing advice are just a few of its many applications. Yet, it entails extensive privacy risk, as trajectory data is extremely privacy-invasive. Unfortunately, several conditions complicate the anonymization of trajectory data: They are sequential, high-dimensional, bound to geophysical restrictions, and easily mapped to semantic POIs.

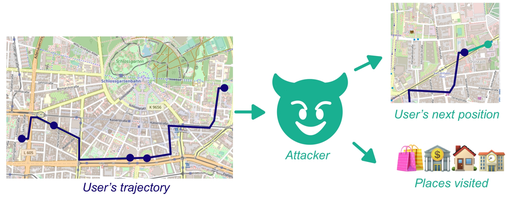


Figure 1. Trajectories may reveal accurate behavioral patterns, allowing attackers to infer sensitive aspects of an individual's life, including health status, religious beliefs, social relationships, or sexual preferences.

Trajectories and Data Sets

Raw Trajectories | Semantic Trajectories

$T = (x_1, y_1, t_1) \rightarrow \cdots \rightarrow (x_n, y_n, t_n)$ — Adds semantic meaning

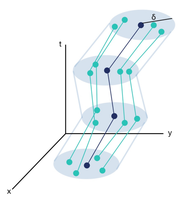
Measuring Privacy

Syntactic Notions

k-anonymity, l-diversity, t-closeness

Semantic Notions

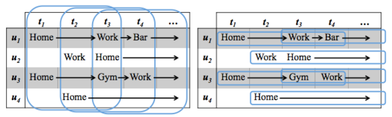
ϵ-differential privacy



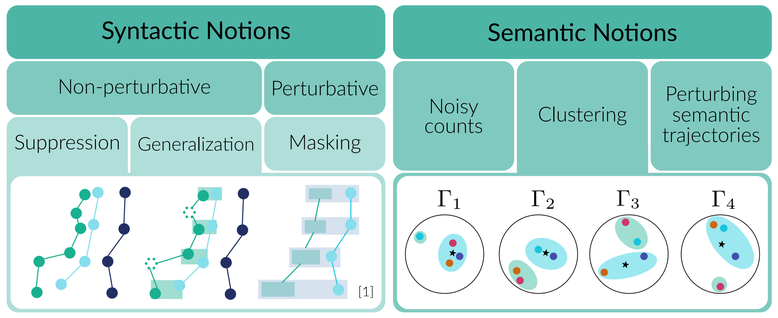
- (k, δ)-anonymity
- k^m-anonymity
- $k^{\tau,\epsilon}$-anonymity
- $(K, C)_L$-privacy
- ...

event-level — user-level

w-event privacy | ℓ-trajectory privacy

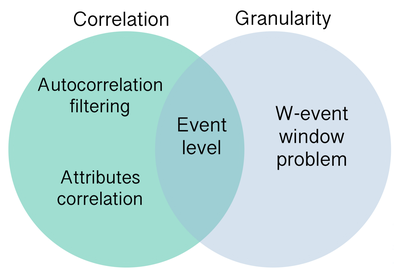


Mechanisms: Achieving Privacy



Privacy Limitations

Shortcomings of semantic notions' privacy guarantees



Limitations in the presented mechanisms

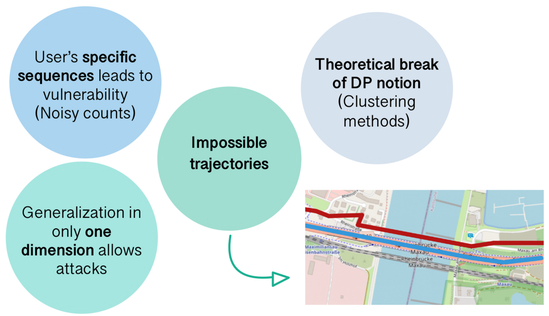


Figure 2. ■ corresponds to an impossible trajectory after sanitation process. With background knowledge (a map), we can rebuild the original trajectory easily.

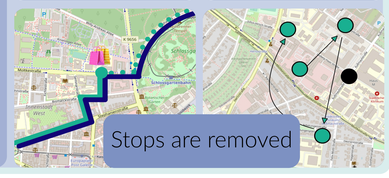
Utility Limitations

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



Opportunities and Future Work

- Development of clustering algorithm for entire trajectories that takes the temporal dimension into account. High-dimensional topological clustering based on persistent homology will be interesting because of its qualitative predominance and its low computational cost.
- Adaptation of alternative notions of privacy that have been proposed to overcome deficiencies of DP against correlations. Creation of new axiomatic notions of privacy and corresponding mechanisms that meet guarantees.
- Consideration of public knowledge to reduce privacy budget consumption and avoid impossible results.
- Revision and selection of proper utility metrics that ensure their applicability.

Acknowledgments & References

J.P.-A. is an Alexander von Humboldt research fellow. The work received support from "la Caixa" Foundation (fellowship code LCF/BQ/PR20/11770009), the EU H2020 programme (Marie Skłodowska-Curie grant agreement No 847648), the Spanish Government (project "COMPROMISE" PID2020-113795RB- C31/AEI/10.13039/501100011033), and the BMBF project "PROPOLIS" (16KIS1393K). The authors at KIT belong to KASTEL SRL (Helmholtz Association Topic 46.23) and Excellence Cluster EXC 2050/1 'CeTI' of Germany's Excellence Strategy (project 390696704).

Map screenshots from © OpenStreetMap contributors [2].

[1] M. E. Nergiz, M. Atzori, Y. Saygin, and B. Güç, "Towards trajectory anonymization: a generalization-based approach," in *SPRINGL '08*, 2008.

[2] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org." https://www.openstreetmap.org, 2022.