On Choosing Between Privacy Preservation Mechanisms for Mobile Trajectory Data Sharing

Rajkarn Singh, George Theodorakopoulos†, Mahesh K. Marina and Myrto Arapinis
The University of Edinburgh, †Cardiff University
Email: r.singh@ed.ac.uk, TheodorakopoulosG@cardiff.ac.uk, mahesh@ed.ac.uk, marapini@inf.ed.ac.uk

I. Overview

The widespread adoption of smart mobile devices equipped with a multitude of sensors offers a rich source of data on user patterns [1]. Open publication of such datasets can be highly valuable for research purposes. However, unscrupulous sharing or publication of such datasets risks violating the privacy of individual mobile users whose data figures in those datasets.

A number of privacy preservation mechanisms have been proposed for trajectory data sharing e.g., differential privacy (DP) [2], k-Anonymity [3] and Plausible Deniability [4]. However, the theoretical guarantees provided by different privacy mechanisms are very different. A clear comparison between existing mechanisms is missing, making it difficult when a data aggregator/owner needs to pick a mechanism for a given application scenario.

In this paper, we aim to quantify privacy offered by different trajectory privacy preservation mechanisms (TPPMs) on a common scale, for the first time, so as to enable comparison across them. Specifically, we propose *STRAP (Scale for TRAjectory Privacy)*, a novel metric that can be used to assess the relative privacy guarantees provided by different TPPMs. We also use STRAP to study how constraining utility affects the privacy achieved by different mechanisms.

Concerning the relevant prior literature, there is very little work on quantifying privacy and the few existing works are limited to specific scenarios and use cases; none of them rigorously compare different state of the art TPPMs.

II. Outline of STRAP Computation

The key idea underlying STRAP is the observation that even though different TPPMs have different theoretical privacy models, all of them perform obfuscation in a way that tries to reduce similarity between the original trajectories and their replications in the output database while also keeping in mind the uniqueness of the users. We use this common ground as a means to compare different mechanisms.

A. Relation between Uniqueness ($\mathcal{U}$) and Privacy

Human mobility traces are known to be highly unique. Hence, we base our mechanism on the uniqueness of records in the original database as uniqueness is a characteristic of the database that captures how much vulnerable database records are to a re-identification attack.

B. Relation between Trajectory Distance ($\mathbb{ED}$) and Privacy

To achieve privacy, TPPMs perform obfuscation by adding noise or by generalization, thereby changing the true values of database attributes. We capture this obfuscation by measuring geographical distance between the actual trajectory of a user and corresponding output trajectories, thus also base our privacy metric on their distance.

C. STRAP Metric Computation

To avoid any bias towards a particular TPPM, our metric design does not rely on the knowledge of mechanism used for privacy preservation, but is only a function of the original database and the output database. STRAP computes the privacy level (Δ_u) for each user trajectory, a_u, as follows [5]:

$$\Delta_u = \frac{\mathbb{ED}(a_u)}{\mathcal{U}_u} \tag{1}$$

Our solution is divided into three main steps as follows:

1) Develop a mobility model that encodes adversary's knowledge input and output trajectories.
2) Using the above mobility model, compute trajectory distance over multiple trajectory segments.
3) Compute trajectory uniqueness and use it as the weight to compute final STRAP value.

III. Evaluation Summary

For evaluation purpose, we selected three state-of-the-art mechanisms based on Differential Privacy (ϵ) [2], k-Anonymity (k_A) [3] and Plausible Deniability (k_{PD}) [4]. Our experiments (details in [5]) show a monotonic increase in privacy (STRAP) with an increase in k_A and k_{PD}, and a decrease in ϵ, thus validating that our privacy metric is in coherence with the various privacy definitions.

To be able to cross-compare different mechanisms, we fix a desired utility level to be achieved and find the STRAP value for each mechanism that provides desired utility. We find that different TPPMs provide very different privacy when we place a utility constraint on them. We observe that when privacy requirements are not stringent, these mechanisms achieve almost similar and high utility values. However, as the utility requirements are relaxed, their performance starts differing from each other in terms of the privacy levels achieved.

References

[1] G. Tsoukaneri et al., "On the inference of user paths from anonymized mobility data," in *IEEE European Sym. on Security and Privacy*, 2016.
[2] X. He, G. Cormode, A. Machanavajjhala, C. M. Procopiuc, and D. Srivastava, "DPT: Differentially private trajectory synthesis using hierarchical reference systems," *Proc. of VLDB Endow.*, 2015.
[3] M. Gramaglia and M. Fiore, "Hiding mobile traffic fingerprints with GLOVE," in *Proceedings of ACM CoNEXT*, 2015.
[4] V. Bindschaedler and R. Shokri, "Synthesizing plausible privacy-preserving location traces," in *IEEE Symposium on Security and Privacy (SP)*, 2016.
[5] R. Singh et.al, "On choosing between privacy preservation mechanisms for mobile trajectory data sharing," in *Proc. IEEE CNS*, 2018.