Towards the publication of privacy-preserving yet useful mobility traces

Vincent Primault
Sonia Ben Mokhtar
Université de Lyon, INSA-Lyon
{firstname.lastname}@liris.cnrs.fr

Cédric Lauradoux
Inria
cedric.lauradoux@inria.fr

Lionel Brunie
Université de Lyon, INSA-Lyon
lionel.brunie@liris.cnrs.fr

Abstract

An increasing amount of mobility data is being collected every day, and privacy is a main concern when publishing this sensitive data. Existing state-of-the-art techniques are often based on adding noise to the geographical data and are not yet fully satisfactory. We propose a novel solution based on time distortion and evaluate it using a real-life dataset.

1. Context

With the widespread adoption of handheld devices, such as smartphones and tablets, more and more mobility data is being gathered every minute. This huge amount of mobility data is very valuable for analysts that can run data mining tasks to analyse users' habits or predict future behaviours. However, mobility data is also very sensitive and its disclosure can be harmful for users' privacy. Attacks that can be found in the literature show that it is not sufficient to simply suppress users' names to anonymise mobility data, though this is a required step. Many attacks aim at extracting users' *points of interests* (POIs). These, are places where users regularly spend some time, like their work place, home or a cinema. These POIs are sensitive as they allow to infer new knowledge such as a user's occupation, his hobbies or even his political or religious preferences. Moreover, by using POIs and some background knowledge it can be possible to guess from "anonymous" traces which trace is his. The literature contains many propositions of protection mechanisms allowing the publication of traces in a privacy-preserving manner. State-of-the art mechanisms often act by adding noise to the location either directly (e.g., [1]) or as a consequence of k-anonymity enforcement (e.g., [2]). This has the side effect of decreasing the utility of published data.

2. Introducing a new protection mechanism

This is why we propose a new protection mechanism. Our goal is to better take into account two conflicting constraints: on one hand there is the necessity to protect users' privacy, on the other hand analysts want useful data from which they can infer accurate observations. Because we believe that spatial information is the most valuable information, we chose to maximize spatial accuracy while allowing a small decrease of temporal accuracy. Our method relies on two complementary steps whose goal is to offset POIs extraction and re-identification attacks. Given the sensitive nature of POIs, we propose to smooth the speed of mobility traces, i.e., we enforce a constant speed along each one. As a consequence, an attacker cannot infer places where users stop, because users seem to be always moving; POIs are hence hidden. However there is still a risk for users to be re-identified by using the shape of their trajectory. This is why in a second step we take advantage of areas where users meet to opportunistically exchange their identities. This makes a re-identification attack more difficult to achieve because there is always a doubt when users' trajectories cross whether or not they have been exchanged.

3. Evaluating our solution

We evaluated the privacy by counting POIs that can still be retrieved after using our solution to anonymise a dataset, w.r.t. to POIs that can be extracted from the original dataset. Results show that under an optimal parametrisation, less than 3 % of POIs can be retrieved, which is very close to or better than our competitors. We evaluated the utility by considering a simple yet useful task which is counting how many users are within a given area during a given time window. With our dataset we found that the relative error of these queries between an anonymized dataset and the original one remains under 22 %, which is better than state-of-the-art mechanisms.

Acknowledgments

This work was supported by the LABEX IMU (ANR-10-LABX-0088) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

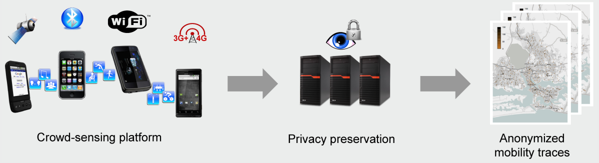
References

[1] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential Privacy for Location-based Systems," in *Proceedings of CCS 2013*. ACM, 2013, pp. 901–914.

[2] O. Abul, F. Bonchi, and M. Nanni, "Anonymization of moving objects databases by clustering and perturbation," *Information Systems*, vol. 35, no. 8, pp. 884–910, 2010.

Towards the publication of privacy-preserving yet useful mobility traces

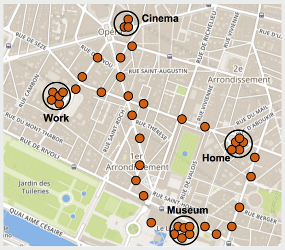
Context



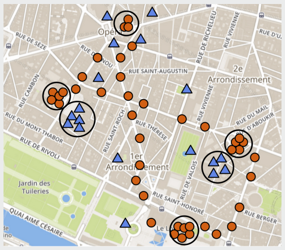
Crowd-sensing platforms are largely used to collect mobility traces. Nonetheless, there are some privacy concerns about mobility traces which are known to be very sensitive and can be harmful for users.

This is why we propose Promesse, a new protection mechanism that enhances location privacy of mobility traces while still allowing analysts to get useful insights from them. We are especially interested in data mining tasks counting users, while maintaining a high spatial accuracy.

Location privacy attacks



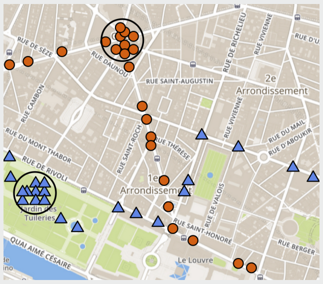
It is possible to extract *points of interest* with clustering algorithms. Then, by attaching semantic labels, they can be used to infer new knowledge about the users.



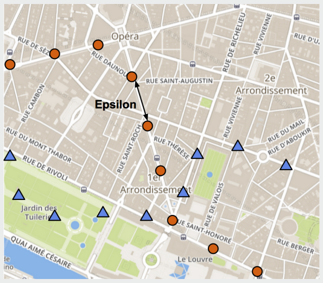
"Are the blue and orange users actually the same person?"

By using points of interest, it can be possible to re-identify physical users.

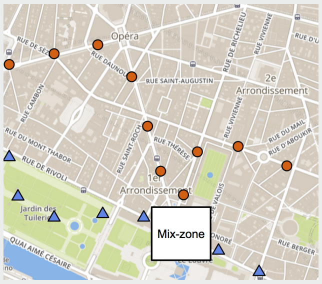
Making the Promesse of a new protection mechanism



Two raw mobility traces, as they would be gathered by a crowd-sensing platform. Each one exhibits one point of interest.



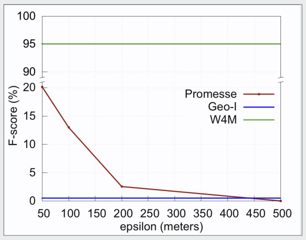
We hide points of interest by enforcing a constant speed along trajectories. This makes a lot more difficult to guess where users actually made a stop.



We take advantage of participants who meet during their day to probabilistically exchange their traces and therefore confuse an attacker trying to re-identify users.

Privacy & utility evaluation

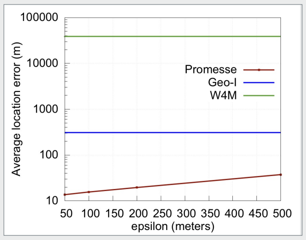
Privacy: hiding points of interest



Metric: We evaluate privacy by computing the F-score of the points of interests that can be extracted from mobility traces before and after anonymization.

Results: If epsilon is big enough, we perform as well as Geo-I and always better than W4M.

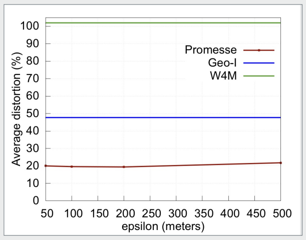
Utility: spatial error



Metric: The average spatial error is the average of the distance between each point of an anonymized trace and the nearest point of the original trace before anonymization.

Results: We keep the average spatial error under 38 meters, well below Geo-I and W4M.

Utility: range queries distortion



Metric: Range queries count how many users are within an area during a time window. The distortion is the relative error between this number of users before and after anonymization.

Results: We maintain the average distortion under 22 %, well below Geo-I and W4M.

Authors

Vincent Primault [1]
vincent.primault@liris.cnrs.fr

Sonia Ben Mokhtar [1]
sonia.ben-mokhtar@liris.cnrs.fr

Cédric Lauradoux [2]
cedric.lauradoux@inria.fr

Lionel Brunie [1]
lionel.brunie@liris.cnrs.fr

[1] Université de Lyon, CNRS, INSA-Lyon
[2] Inria