DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics

Collecting, analyzing and publishing location / mobility data without raising privacy concerns through decentralized analysis and storage.

Simon van Endern

DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics

Collecting, analyzing and publishing location / mobility data without raising privacy concerns through decentralized analysis and storage.

Titel der Abschlussarbeit auf Deutsch

Author:	Simon van Endern
Supervisor:	Prof. Dr.-Ing. Jörg Ott
Advisor:	Trinh Viet Doan
Submission Date:	30.06.2019

I confirm that this bachelor's thesis in informatics is my own work and I have documented all sources and material used.

Munich, 30.06.2019 Simon van Endern

Acknowledgments

Abstract

We propose a method to publish location ata without raising privacy concerns.

Contents

Acknowledgments **iii**

Abstract **iv**

1 Introduction **1**
1.1 Existing approaches 2
1.2 Section 2
1.2.1 Subsection 2

2 Method **4**

3 Solution **5**

4 Analysis **6**

5 Conslusion **7**

List of Figures **8**

List of Tables **9**

Bibliography **10**

1 Introduction

The introduction is meant to motivate the subject area (why is this important?), define the problem you are interested in (what are you doing?), and limit the scope (where do you stop?). It also gives an outline of the thesis (which chapters will explain what?) and explains how you are going the approach your subject.

With the advent of the internet and large-scale applications, the question of privacy has drawn increasing attention. Especially with services like Twitter, Facebook, Google & Co. there are problems and privacy infringements when user data is realeased. One example is that the location data of Twitter tweets was published without asking the user for permission. Furthermore this data is only available through the API, so that the user is not aware of this infringement. Using this data, [ZB11] has shown that this data can be used to infer a users home address and often also the work address, even if the user itself is privacy-aware, thus does not publish his / her name, etc.

[Dra+19] finds that even when personal data is anonymized thus that names and addresses, etc. are removed, sensitive information can be inferred from the data. In this study it was shown that from call-records in the US the home address and also often the work address of a person could be inferred. They highlight that while adhering to the k-anomymity model proposed by [Swe02] it is practically not possible to publish datasets that are still of any significant use.

We will use the definition of location privacy as defined by [BS03]: " the ability to prevent other parties from learning one's current or past location". They further propose a different approach to preserve privacy. TODO!!!

Also [GP09] highlights the thread that home and work locations can be inferred from anonymized datasets and can in combination with other sources yield even more information about a user. To reduce this risk, they propose "to collect the minimum amount of information needed". In contrary, we want to investigate another approach, so that rich data can still be used and be published in an aggregated manner to let people profit from the data but still preserve privacy.

This research shows that publishing raw data is critical, even when the data is anonymized. As still this data could be useful for many stakeholders, we will investigate how on the one hand aggregated datacan be published without imposing any privacy

risk to the owners of the data and on the other hand develop a prototype of a mobile application through which this location data is aggregated in a decentralized manner so that the raw user data never leaves the users' device.

Another problem that arises is that anonymization algorithms applied to datasets prior to publishing them might yield good results if the location data is in a densily populated area but might perform poorly if the population is only sparse [Hoh+07].

1.1 Existing approaches

- Collect less data [GP09]
- Mixing approach [BS03]
- Anonymize data to meet the kriteria of k-anonymity [Swe02] and [Dra+19]
- spatial cloaking [Kru07]

1.2 Section

Citation test [Lam94].

1.2.1 Subsection

See Table 1.1, Figure 1.1, Figure 1.2, Figure 1.3.

Table 1.1: An example for a simple table.

A	B	C	D
1	2	1	2
2	3	2	3

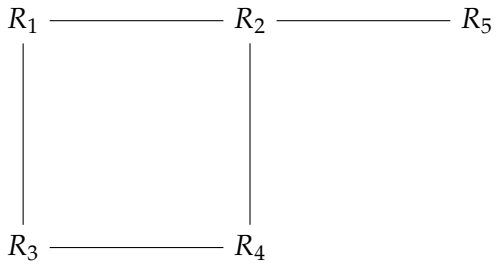

Figure 1.1: An example for a simple drawing.

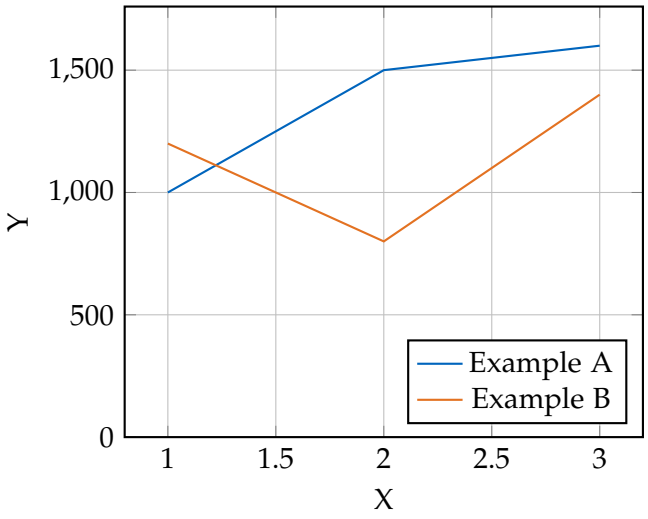

Figure 1.2: An example for a simple plot.

```
SELECT * FROM tbl WHERE tbl.str = "str"
```

Figure 1.3: An example for a source code listing.

2 Method

3 Solution

4 Analysis

5 Conslusion

List of Figures

1.1 Example drawing 3
1.2 Example plot 3
1.3 Example listing 3

List of Tables

1.1 Example table 2

Bibliography

[BS03] A. R. Beresford and F. Stajano. "Location privacy in pervasive computing." In: *IEEE Pervasive computing* 1 (2003), pp. 46–55.

[Dra+19] K. Drakonakis, P. Ilia, S. Ioannidis, and J. Polakis. "Please Forget Where I Was Last Summer: The Privacy Risks of Public Location (Meta)Data." In: *CoRR* abs/1901.00897 (2019). arXiv: 1901.00897.

[GP09] P. Golle and K. Partridge. "On the Anonymity of Home/Work Location Pairs." In: *Proceedings of the 7th International Conference on Pervasive Computing*. Pervasive '09. Nara, Japan: Springer-Verlag, 2009, pp. 390–397. ISBN: 978-3-642-01515-1. DOI: 10.1007/978-3-642-01516-8_26.

[Hoh+07] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. "Preserving Privacy in Gps Traces via Uncertainty-aware Path Cloaking." In: *Proceedings of the 14th ACM Conference on Computer and Communications Security*. CCS '07. Alexandria, Virginia, USA: ACM, 2007, pp. 161–171. ISBN: 978-1-59593-703-2. DOI: 10.1145/1315245.1315266.

[Kru07] J. Krumm. "Inference Attacks on Location Tracks." In: *Pervasive Computing*. Ed. by A. LaMarca, M. Langheinrich, and K. N. Truong. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 127–143. ISBN: 978-3-540-72037-9.

[Lam94] L. Lamport. *LaTeX : A Documentation Preparation System User's Guide and Reference Manual*. Addison-Wesley Professional, 1994.

[Swe02] L. Sweeney. "k-anonymity: A model for protecting privacy." In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10.05 (2002), pp. 557–570.

[ZB11] H. Zang and J. Bolot. "Anonymization of location data does not work: A large-scale measurement study." In: *Proceedings of the 17th annual international conference on Mobile computing and networking*. ACM. 2011, pp. 145–156.