Methods for Location Privacy: A comparative overview

Kostantinos Chatzikokolakis
CNRS, École Polytechnique,
University of Paris Saclay, France

Ehab ElSalamouny
Faculty of Computers and Informatics,
Suez Canal University, Egypt

Catuscia Palamidessi
INRIA,
University of Paris Saclay, France

Anna Pazii
INRIA,
University of Paris Saclay, France

the essence of knowledge
Boston — Delft

Foundations and Trends® in Privacy and Security

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

K. Chatzikokolakis, E. ElSalamouny, C. Palamidessi and A. Pazii. *Methods for Location Privacy: A comparative overview*. Foundations and Trends® in Privacy and Security, vol. 1, no. 4, pp. 199–257, 2017.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-68083-366-9

Foundations and Trends® in Privacy and Security

Volume 1, Issue 4, 2017

Editorial Scope

Topics

Foundations and Trends® in Privacy and Security publishes survey and tutorial articles in the following topics:

- Access control
- Accountability
- Anonymity
- Application security
- Artifical intelligence methods in security and privacy
- Authentication
- Big data analytics and privacy
- Cloud security
- Cyber-physical systems security and privacy
- Distributed systems security and privacy
- Embedded systems security and privacy
- Forensics
- Hardware security
- Human factors in security and privacy
- Information flow
- Intrusion detection
- Malware
- Metrics
- Mobile security and privacy
- Language-based security and privacy
- Network security
- Privacy-preserving systems
- Protocol security
- Security and privacy policies
- Security architectures
- System security
- Web security and privacy

Information for Librarians

Foundations and Trends® in Privacy and Security, 2017, Volume 1, 4 issues. ISSN paper version pending. ISSN online version pending. Also available as a combined paper and online subscription.

Foundations and Trends® in Privacy and Security
Vol. 1, No. 4 (2017) 199–257

DOI: 10.1561/3300000017

Methods for Location Privacy: A comparative overview

Kostantinos Chatzikokolakis
CNRS, École Polytechnique,
University of Paris Saclay, France

Ehab ElSalamouny
Faculty of Computers and Informatics,
Suez Canal University, Egypt

Catuscia Palamidessi
INRIA,
University of Paris Saclay, France

Anna Pazii
INRIA,
University of Paris Saclay, France

Contents

Abstract

The growing popularity of Location-Based Services, allowing for the collection of huge amounts of information regarding users' locations, has started raising serious privacy concerns. In this survey we analyze the various kinds of privacy breaches that may arise in connection with the use of location-based services, and we consider and compare some of the mechanisms and the metrics that have been proposed to protect the user's privacy, focusing in particular on the comparison between probabilistic spatial obfuscation techniques.

K. Chatzikokolakis, E. ElSalamouny, C. Palamidessi and A. Pazii. *Methods for Location Privacy: A comparative overview*. Foundations and Trends® in Privacy and Security, vol. 1, no. 4, pp. 199–257, 2017.
DOI: 10.1561/3300000017.

1

The problems of privacy in location-based services

In recent years, the growing popularity of mobile devices equipped with GPS chips, in combination with the increasing availability of wireless data connections, has led to a growing use of Location-Based Services (LBSs), namely applications in which a user obtains, typically in real-time, a service related to his current location. Recent studies of the Pew Research Center show that in 2017, 77% of the adult population of the US owns a smartphone (in comparison with 35% in 2011) [63], and according to the same institution's last survey about LBSs, in 2013, a high percentage (74%) of the smartphone owners used services based on their location [99]. Examples of LBSs include mapping applications (e.g. Google Maps), Points of Interest (POI) retrieval (e.g. AroundMe), coupon/discount providers (e.g. GroupOn) and location-aware social networks (e.g. Foursquare).

LBS providers often collect and store users' locations and mobility traces (sequences of spatio-temporal points representing the users' itineraries), for the purpose of further utilization, possibly by a third-party. For instance, they can be used for statistical analyses, such as finding typical mobility patterns and popular places [74, 97]), or they can be made public to provide additional services to users, such as traffic information [44].

While LBSs have demonstrated to provide enormous benefits to individuals and society, the growing exposure of users' location information raises important privacy issues. Not only the experts, but also the population at large are becoming increasingly aware of the risks, due to the repeated cases of violations and leaks that keep appearing on the news. For instance, on April 20th, 2011 it was discovered that the iPhone was storing and collecting location data about the user, syncing them with iTunes and transmitting them to Apple, all without the user's knowledge. More recently, the Guardian has revealed, on the basis of the documents provided by Edward Snowden, that the NSA and the GCHQ have been using certain smartphone apps, such as the wildly popular Angry Birds game, to collect users' private information such as age, gender and location [6], again without the users' knowledge. Another case regards the Tinder application, which was found sharing the exact latitude and longitude co-ordinates of users as well as their birth dates and Facebook IDs [73]; even after the initial problem was fixed, it was still sharing more accurate location data than intended, as users could be located to within 100 feet of their present location [26].

A major source of concern about location privacy lies in the realization that with sufficiently accurate data, it is possible to precisely locate a user and track his movements throughout the day [18], giving rise to a variety of malicious activities such as robbing or stalking. For instance, in Wisconsin there were episodes of men tracking women with GPS or other location devices [60]. In California, records from automatic toll booths on bridges were used in divorce proceedings to prove claims about suspicious movements of spouses [82]. The application "Girls Around Me", combined social media and location information to find nearby women who did not necessarily agree to be found, allowing to access their Facebook profiles with a single click [11]. Particularly worrisome is the perspective of potential combination with the users' most sensitive information, such as sexual orientation.

To some extent, the research and the experimentation on privacy contribute to raise the awareness about the practical risks. For instance, the website "Please Rob Me" [65] aggregates location check-ins and presents them as “robbery opportunities”, pointing out the fact that publicly announcing one’s location effectively reveals to the world that they are not home.

1.1 Classification of threats

Following [35], we classify the concerns about the leakage of location information into three major kinds of threats:

Tracking Threat: An adversary collecting continuously the location updates of the user might be able to identify the user’s mobility patterns (frequently traveled routes) and predict his present and future location with high accuracy by leveraging typical mobility habits [47, 94].

Identification Threat: The adversary can use the user’s traces as quasi-identifiers to reveal his identity in an anonymized dataset. This may happen even if the adversary accesses the user’s location only sporadically, since he might be able to infer his frequently visited locations, such as home and work. This is the most studied kind of threat in the literature, we expand on it in the next section.

Profiling Threat: Mobility traces, and in particular the points of interest that can be extracted from them, typically contain semantic information that the adversary can use for *profiling*, that is for infering a variety of (often sensitive) information about the user. Examples include health clinics, religious places, areas which may revel his sexual inclinations, etc. [5]. The practice of location profiling is likely to increase in the future, as marketers are becoming more and more aware of its potential to gain visibility of consumer behavior in the real world, and to help targeting their marketing efforts. Indeed, location profiling seems to provide insights into offline activity at a level comparable to that of web or mobile app analytics for online activity. There are already various companies that provide this kind of services: for instance, Urban Airship [89] offers tools that produce audience profiles by combining in-app behaviors, user preferences, and location. Mobility data are particularly useful, since brands can segment users based on their current or past location.

1.2 Identification of the user from his traces

In this section we focus on the threat constituted by using location data for fingerprinting the user, namely for finding out the identity of the person who has originated the data. In short, the problem raises by the fact that mobility traces may be *unique* to an individual, and they can therefore allow identifying that individual like the ridges on his finger. Apart from uniqueness, *temporal correlation* is also crucial for fingerprinting, allowing an anonymized trace to be identified based on mobility data about the same individual that have been previously recorded.

1.2.1 Uniqueness of human mobility traces.

There have been various statistical studies aimed at showing the uniqueness of human mobility traces. One of the most remarkable ones is that of de Montjoye et al. [23], measuring uniqueness in the following way. Given a set of points P, and a set of traces T, we say that P identifies a unique trace in T if there is exactly one trace in T that contains P. Then, the uniqueness of T is defined as the percentage of traces in T that are uniquely identified by a set of n points drawn randomly from a random trace in T (where n is a parameter). They examined fifteen months of human mobility traces generated by 1.5 million of individuals, who were users of a certain mobile phone operator. Each time a user interacted with the network by initiating or receiving a call or a text message, the location of the connecting antenna was recorded in the dataset together with the time of the event, and linked to previous location-time points of the same user already in the dataset, via the user id, so to form a trace (one trace for each user). The experiments showed that human mobility traces are highly unique: In fact, with the temporal granularity fixed to an hour, and the spatial granularity equal to that given by the carrier's antennas, 4 spatio-temporal points, randomly drawn from a trace, were enough to uniquely identify the trace in 95% of the cases. They also observed that the uniqueness of mobility traces decays approximately as the $1/10$ power of the spatial and temporal resolution. Hence, they concluded that even coarse datasets provide little anonymity.

Song et al. [83] conducted similar experiments on a dataset of location-time data generated by about a million users over a period of a week. They considered the same notion of unique identification as de Montjoye et al., except that they calculated the percentage of identification on all the traces instead than some randomly drawn subset. The location of each individual was recorded every fifteen minutes. The spatial resolution of the data (i.e., the minimum distance between two locations) was about 0.11 km, while the diameter of the whole area (i.e., the largest distance between two locations) was about 49 km. Their results confirm that, even with a low resolution, location traces can be identified with only a few spatio-temporal points. In particular, they show that 2 points are enough to uniquely identify a trace in 60% of the cases.

It is important to note that the implicit notion of attack considered in the above works presupposes that the adversary is provided with points that he had "previously seen" in a trace, and the only challenge (for the adversary) is to be able to distinguish which trace. In contrast, Rossi et al. [68] considered the threat posed by a "previously unseen" set of points. Namely, they assume that the attacker has already collected a set of traces T from some community of users, one trace per user, and then, given a set of additional points P produced by one of the users during his trajectory, they try to re-identify the user by looking for the closest trace, namely the trace in T with the smallest Hausdorff distance from P. They experimented with three real-world datasets GPS mobility traces: CabSpotting [64][1], CenceMe [59] and GeoLife [55]. The location data in these datasets have high spatial resolution (GPS coordinates up to 5 or 6 digits precision). As for the temporal resolution, in GeoLife and CabSpotting locations are recorded

[1]Althoug [68] refers to CabSpotting, the citation is relative to a mobility traces dataset called CRAWDAD.

at a time interval of $1-5$ seconds, while for CenceMe it is 1 hour. Concerning the experiments methodology, they randomly partitioned each dataset into a *training set* and a *test set*, where each trace contained 50% of the original GPS points. Then, they used the training set as the traces T to identify, using sets of points P extracted randomly from traces in the test set. They showed that, thanks to the high precision of the GPS coordinates, on GeoLife and CenceMe just 1 spatio-temporal point is enough to identify 90% and 96% of the traces, respectively. With 2 points, these percentages reach 94% and 99%. The results for CabSpotting are significantly lower: 60% for 2 points. The difference is probably due to the nature of the data: GeoLife and CenceMe contain traces left by users during their daily routines, while CabSpotting are traces of taxi drivers in the San Francisco Bay area. The first two contain many personal and thus unique locations, such as home and workplace locations, while the latter is characterized by the presence of common taxi routes and locations associated to taxi ranks.

1.2.2 Reconstructing traces from location samples

Typically, there can be various users repeatedly updating and sending their positions on the map to some LBS. Hence, collecting these locations may result in a mix-up of traces left by different individuals. Un-mixing the locations, i.e., reconstructing the individual traces, can be done easily when the data are associated to some invariant attribute, like, for instance, a pseudonym. Even when the data are completely anonymous, however, the traces can often still be reconstructed by linking the location samples. Clearly, the higher is the sample frequency compared to the users' density in the area, the easier it is to recognize a trace (cfr. Figure 1.1). In fact, the next point in a trajectory will be at a distance determined by the speed of the user and the time in between the two updates. The reconstruction of a trace can also be facilitated by correlating location samples with likely routes on a map. Finally, the task can be enhanced by using a model of typical trajectories constructed on the basis of prior observations on the population movements.

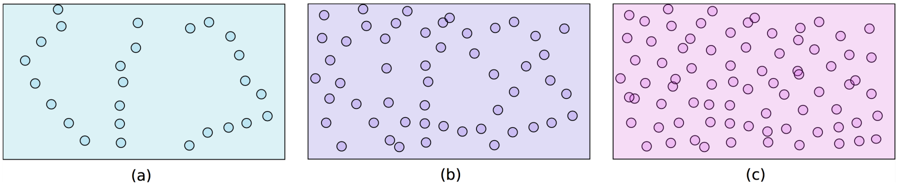

Figure 1.1: Traces in a low (a), medium (b), and high density area (c)

The first attempt to reconstruct the traces from completely anonymized mobility data (i.e., without any pseudonyms) was by Gruteser et al. [42]. They used a multi-tracking algorithm to identify individual mobility traces from a collection of anonymized location samples generated by multiple users. They tested their algorithm on a collection of GPS traces generated by the students of a university campus, and their experiments showed that often individuals used to travel along the same unique route and could therefore be re-identified. Their system however was prone to misclassification of crossing paths, as it was unable to determine whether the paths of different individuals actually crossed or just touched.

More recently, Tsoukaneri et al. [87] developed a mechanism called *Comber* which is able to disentangle the traces by using a generic, empirically derived histogram of user speeds. The authors evaluated Comber with two different datasets, MDC [45] and GeoLife [55], which consist both of GPS-based mobility traces (collected in Lausanne and Beijing, respectively). Each of these datasets span more than a year and include location information of about 180 users. Their results show that Comber is able to infer the original traces of the users with more than 90% accuracy.

1.2.3 Linking traces to users' identity

There has been a lot of research showing that it is possible to infer user identities from anonymous traces, especially when the traces are pseudonymized (i.e. the real identity has been replaced by a pseudonym) rather than completely anonymized. Beresford and Stajano [8] already pointed out that the re-identification risks of LBS' users employing pseudonyms: they showed that almost all location traces of AT&T Labs Cambridge employees collected from the Active Bat system could be correctly identified by knowing the office positions of the workers and by keeping track of the frequency of visits of a given pseudonym to each office.

Many of the attacks on pseudonymized traces are, like the above, based on observing the frequent presence of the pseudonyms in specific locations that can be easily linked to a certain individual, like home or office. For instance, Krumm [48] proposed various algorithms to infer the user's home address, and used a web search engine in order to reveal the real identities of the subjects. Notably, Golle and Partridge [40], using US census data, showed that knowing both locations of an individual's home and workplace with the precision of a census block allowed to uniquely identify most of the U.S. working population. Furthermore, even with the lower granularity of a census track, although the average size of the anonymity set (i.e., the number of people sharing the same pair) went up to 21, the location data of people who lived and worked in different regions could still be easily re-identified.

A further study [96] investigated call records rather than census data, using a data set of more than 30 billion call records made by 25 million cell phone users in the US. They considered the "top N" locations for each user, inferred from the call records, and different levels of granularity, ranging from a cell sector to whole cell (where cell and cell sector are location units used by the phone company) to the zip code, city, county and state. They analysed a variety of different factors potentially impacting the size of the anonymity set, such as the distance between the top N locations, the geographic environment (rural vs urban), and social information (whether the size of the user's social network is large or small). Their result showed that, while the top 1 location does not typically yield small anonymity sets, the top 2 and top 3 locations do, at least at the sector or cell-level granularity. For example, with top 3 locations, 85% of the users are identifiable at the sector level, 50% at the cell level, and 35% at the zip code level.

Even when the location data are completely anonymized (i.e., no pseudonym is used), though, it is still possible to retrieve the user's identity by means of modern machine learning technologies if the attacker disposes of side information about the user. Several works in the literature have investigated this problem, particularly in the case in which a database of users' profiles in the form of previously collected traces, called *the training set*, is available to the adversary. The work by Rossi et al. [68] mentioned in § 1.2.1 went in this direction; however it did not use the full power of machine learning techniques, and it was more focused in the uniqueness of traces rather than re-identification of the user. In general, the idea is that the adversary will use the training set to build a representation of the users' typical movements. Thus each user will be associated to a mathematical model of his past traces, playing the role of a signature. This model can be, for instance, a Markov chain, but other models have been investigated as well. Then the attacker will collect one or more of the victim's (sanitized) traces, *the testing set*, from which he will build a model as well. The latter is then compared to the models of the training set, according to some similarity criterion, and the user profile most likely to correspond to the target user is finally selected.

De Mulder et al. [24] investigated this kind of attack on mobility traces generated by a GSM cellular network. They developed two methods based on different models and on the cosine similarity measure, and evaluated them on the Reality Mining dataset made available by the MIT Media Lab, which consists of the location traces of one hundred human subjects at MIT during the 2004 − 2005 academic year, collected using one hundred instrumented Nokia 6600 smart phones. With the best of those methods, they were able to re-identify about 80% of the users. It is to be noted that a trace generated by a GSM network is formed by the sequence of all cells that the user has visited along his path, i.e., it is not possible to skip cells by "jumping" to a non-adjacent cell. This may affect the success rate when compared with the case in which the traces consist of locations generated dynamically with, say, a GPS.

Ma et al. [52] considered also two kinds of adversaries: passive ones, retrieving the testing set from a public source, and active ones that can deliberately participate or perturb the data collection phase to gain additional knowledge. The authors used four different estimators to measure the similarity between mobility traces: the Maximum Likelihood Estimator, relying on the Euclidean distance, the Minimum Square Approach, computing the sum of the square of the difference between the traces, the Basic Approach, which assumes that the traces might been perturbed by uniform noise, and the Weighted Exponential Approach, which is similar to the previous one except that no assumption is made on the type of noise generated. The authors tested their methods on two datasets: the CRAWDAD repository [64], recording the movements of San Francisco YellowCabs, and a collection of traces generated by the public buses in Shanghai city. They obtained a success rate of deanonymization of 80% to 90%, even in the presence of noise.

Both [52] and [24], however, took the samples to generate the testing set directly from the training set. Clearly such way of proceeding introduced a bias that may have resulted in an overly strong success rate in the re-identification results. In fact Gambs et al. [36] showed that there is a substantial difference in the success rate when the training set and the testing set are separated. They used a model based on Mobility Markov Chains, namely Markov chains where the states are locations. They considered various similarity measures between such chains, and tested their methods on several GPS datasets, including MDC and Geolife. For each individual, they split his mobility traces, chronologically ordered, into two disjoint parts of approximately the same size: the first half formed the training set, and the second half the testing set. Thus the training and the testing data were not only disjoint, but also separated in time. With such split, they were able to re-identify between 35% and 45% or the users. For comparison, they repeated the experiments also without splitting, i.e., using the same set of traces for training and for testing, and obtained, in this case, a success rate of almost 100%! Of course, this comparison is not completely fair because they used as testing set exactly the same as the training set, instead than a subset as in previous works. Nevertheless, such high success rate shows that (1) the training set and the testing set should be independent to avoid any bias, and (2) the Mobility Markov Chain obtained from the traces of a user is almost always unique to the user.

1.3 The users' point of view

The users' concerns about location privacy, and privacy in general, vary a lot from individual to individual, and depend on factors such as age, education, cultural background, etc. They also tend to evolve in time, and cases of privacy breaches that hit the news, like that of "Birds and 'leaky' phone apps" [6], can have a huge impact on the attitude of the population.

There have been several studies to assess people's perceptions and attitude towards privacy. We mention in particular the empirical research conducted at CMU by Acquisti and his team, which provides a systematic analysis of several aspects of human behavior in relation to privacy. See [1] for a summary of their findings.

Concerning the specific case of location privacy, the concerns seem in general less strong than for other kinds of sensitive data (such as medical records, financial data, bank information etc.), and the studies give mixed results. For instance, in 2014 the authors of [35] interviewed 180 smartphone users, recruited through social network announcements and through Amazon Mechanical Turk. They chose Mechanical Turk workers who had achieved master qualification. They obtained the following statistics: 78% of the participants believed that apps accessing their location can pose privacy threats. Also, 85% of them reported that they care about who accesses their location information (in line with the 87% reported by the 2011 Microsoft survey [56]). Furthermore, 77% of the users were interested in installing a privacy protection mechanism. Finally, on the specific method based on the addition of random noise, 52% of the surveyed individuals stated no problem in supplying apps with imprecise location information to protect their privacy. Only 18% of the surveyed people objected to supplying apps with imprecise location information.

On the other hand, in contrast with the other kinds of sensitive data mentioned above (medical record etc.) there seem to be more willingness to renounce to location privacy in exchange of compensation. For instance, Danezis et al. [22] conducted a study on 74 undergraduates to find how much money they would require in order to share a month's worth of their location data. The median price was £10 if the data were to be used for research purposes, and £20 if the data were to be used commercially. In [49] the author says that he could we easily convince over 250 people from his institution to give him two weeks of GPS data recorded in their car in return for a 1% chance of winning a US$ 200 MP3 player. He asked 97 of them if he could share their location data outside our institution, and only 20% said 'no'. In contrast, in an experiment conducted by Acquisti et al. [2] on the privacy attitude towards payments, where people were offered the choice between a traceable gift card of 12 US$ or an anonymous gift card of 10 US$, about half of the people chose the second option. Incidentally, [2] main point is to show that people value their privacy differently, depending on how the choice privacy vs non-privacy is presented to them. In particular, people tend to assign a different value to their privacy depending on whether they would receive a compensation in order to disclose otherwise private information, or rather they would pay to protect otherwise public information.

In conclusion, location data seems to be less critical in the mind of many people than data like financial or medical ones, but this may be due to the lack of knowledge about the negative consequences of a location leak. In particular, about the fact that the location can help profiling the user with respect to more sensitive data. Furthermore, the attitude of people concerning the protection of location information may change during time, along with the general increase of privacy concerns. For example, a study in [1] showed that, in the last decade, the percentage of members in the Carnegie Mellon University Facebook network who chose to publicly reveal personal information had decreased steadily. For instance, over 80% of profiles publicly revealed their birthday in 2005, but less than 20% in 2011.

References

[1] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. Privacy and human behavior in the age of information. *Science*, 347(6221):509–514, 2015.

[2] Alessandro Acquisti, Leslie K. John, and George Loewenstein. What is privacy worth? *The Journal of Legal Studies*, 42(2):249 – 274, 2013.

[3] Miguel E. Andrés, Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Geo-indistinguishability: differential privacy for location-based systems. In *Proceedings of the 20th ACM Conference on Computer and Communications Security (CCS 2013)*, pages 901–914. ACM, 2013.

[4] Claudio Agostino Ardagna, Marco Cremonini, Ernesto Damiani, Sabrina De Capitani di Vimercati, and Pierangela Samarati. Location privacy protection through obfuscation-based techniques. In Steve Barker and Gail-Joon Ahn, editors, *Proc. of the 21st Annual IFIP WG 11.3 Working Conference on Data and Applications Security (DAS)*, volume 4602 of *Lecture Notes in Computer Science*, pages 47–60. Springer, 2007.

[5] Daniel Ashbrook and Thad Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, 7(5):275–286, 2003.

[6] James Ball. Angry birds and 'leaky' phone apps targeted by NSA and GCHQ for user data. The Guardian, January 27, 2014. `http://www.theguardian.com/world/2014/jan/27/nsa-gchq-smartphone-app-angry-birds-personal-data`.

[7] Bhuvan Bamba, Ling Liu, Péter Pesti, and Ting Wang. Supporting anonymous location queries in mobile environments with privacygrid. In *Proc. of the 17th International Conference on World Wide Web (WWW)*, pages 237–246. ACM, 2008.

[8] Alastair R. Beresford and Frank Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.

[9] Claudio Bettini, Xiaoyang Sean Wang, and Sushil Jajodia. Protecting privacy against location-based personal identification. In *Proceeding of the 2nd Workshop on Secure Data Management (SDM 2005)*. Springer, 2005.

[10] Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Optimal geo-indistinguishable mechanisms for location privacy. In *Proceedings of the 21th ACM Conference on Computer and Communications Security (CCS 2014)*, 2014.

[11] J. Brownlee. This Creepy App Isn't Just Stalking Women Without Their Knowledge, It's A Wake-Up Call About Facebook Privacy (Update), March 2012. http://www.cultofmac.com/157641/.

[12] Konstantinos Chatzikokolakis, Miguel E. Andrés, Nicolás E. Bordenabe, and Catuscia Palamidessi. Broadening the scope of Differential Privacy using metrics. In Emiliano De Cristofaro and Matthew Wright, editors, *Proceedings of the 13th International Symposium on Privacy Enhancing Technologies (PETS 2013)*, volume 7981 of *Lecture Notes in Computer Science*, pages 82–102. Springer, 2013.

[13] Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Marco Stronati. A predictive differentially-private mechanism for mobility traces. In E. De Cristofaro and S.J. Murdoch, editors, *Proceedings of the 14th International Symposium on Privacy Enhancing Technologies (PETS 2014)*, volume 8555 of *Lecture Notes in Computer Science*, pages 21–41. Springer, 2014.

[14] Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Marco Stronati. Constructing elastic distinguishability metrics for location privacy. *PoPETs*, 2015(2):156–170, 2015.

[15] Kostantinos Chatzikokolakis, Ehab ElSalamouny, and Catuscia Palamidessi. Efficient utility improvement for location privacy. *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2017(4):308–328, 2017.

[16] Rui Chen, Gergely Ács, and Claude Castelluccia. Differentially private sequential data publication via variable-length n-grams. In Ting Yu, George Danezis, and Virgil D. Gligor, editors, *Proceedings of the 19th ACM Conference on Computer and Communications Security (CCS 2012)*, pages 638–649. ACM, 2012.

[17] Reynold Cheng, Yu Zhang, Elisa Bertino, and Sunil Prabhakar. Preserving user location privacy in mobile data management infrastructures. In George Danezis and Philippe Golle, editors, *Proceedings of the 6th International Workshop on Privacy Enhancing Technologies (PET)*, volume 4258 of *Lecture Notes in Computer Science*, pages 393–412. Springer, 2006.

[18] Anne Cheung. Location privacy: The challenges of mobile service devices. *Computer Law & Security Review*, 30(1):41–54, 2014.

[19] Chi-Yin Chow. Cloaking algorithms for location privacy. In Shashi Shekhar, Hui Xiong, and Xun Zhou, editors, *Encyclopedia of GIS.*, pages 229–235. Springer, 2017.

[20] Chi-Yin Chow and Mohamed F. Mokbel. Trajectory privacy in location-based services and data publication. *SIGKDD Explorations*, 13(1):19–29, 2011.

[21] Chi-Yin Chow, Mohamed F. Mokbel, and Xuan Liu. Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments. *Geoinformatica*, 15(2):351–380, April 2011.

[22] George Danezis, Stephen Lewis, and Ross J. Anderson. How much is location privacy worth? In *Proceedings of the 4th Annual Workshop on the Economics of Information Security, (WEIS 2005)*, 2005.

[23] Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. Unique in the crowd: The privacy bounds of human mobility. *Nature Scientific Reports*, 3(1376), 03 2013.

[24] Yoni De Mulder, George Danezis, Lejla Batina, and Bart Preneel. Identification via location-profiling in gsm networks. In *Proceedings of the 7th ACM Workshop on Privacy in the Electronic Society (WPES 2008)*, pages 23–32. ACM, 2008.

[25] Rinku Dewri. Local differential perturbations: Location privacy under approximate knowledge attackers. *IEEE Transactions on Mobile Computing*, 99(PrePrints):1, 2012.

[26] Stuart Dredge. Tinder dating app was sharing more of users' location data than they realised. The Guardian, February 2014. `https://www.theguardian.com/technology/2014/feb/20/tinder-app-dating-data-location-sharing`.

[27] Matt Duckham and Lars Kulik. A formal model of obfuscation and negotiation for location privacy. In *Proc. of the Third International Conference on Pervasive Computing (PERVASIVE)*, volume 3468 of *Lecture Notes in Computer Science*, pages 152–170. Springer, 2005.

[28] Cynthia Dwork. Differential privacy. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *33rd International Colloquium on Automata, Languages and Programming (ICALP 2006)*, volume 4052 of *Lecture Notes in Computer Science*, pages 1–12. Springer, 2006.

[29] Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In Michael Mitzenmacher, editor, *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC)*, pages 371–380. ACM, May 31 - June 2 2009.

[30] Cynthia Dwork, Frank Mcsherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *In Proceedings of the Third Theory of Cryptography Conference (TCC)*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer, 2006.

[31] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[32] Ehab ElSalamouny, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Generalized differential privacy: Regions of priors that admit robust optimal mechanisms. In Franck van Breugel, Elham Kashefi, Catuscia Palamidessi, and Jan Rutten, editors, *Horizons of the Mind. A Tribute to Prakash Panangaden*, volume 8464 of *Lecture Notes in Computer Science*, pages 292–318. Springer International Publishing, 2014.

[33] Ehab ElSalamouny and Sébastien Gambs. Optimal noise functions for location privacy on continuous regions. *International Journal of Information Security*, 2017.

[34] Kassem Fawaz, Huan Feng, and Kang G. Shin. Anatomization and protection of mobile apps' location privacy threats. In Jaeyeon Jung and Thorsten Holz, editors, *Proceedings of the 24th USENIX Security Symposium, (USENIX Security 2015)*, pages 753–768. USENIX Association, 2015.

[35] Kassem Fawaz and Kang G. Shin. Location privacy protection for smartphone users. In *Proceedings of the 21st ACM Conference on Computer and Communications Security (CCS 2014)*, pages 239–250. ACM Press, 2014.

[36] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. De-anonymization attack on geolocated data. *J. Comput. Syst. Sci.*, 80(8):1597–1614, 2014.

[37] Bugra Gedik and Ling Liu. Location privacy in mobile systems: A personalized anonymization model. In *Proc. of the 25th International Conference on Distributed Computing Systems (ICDCS)*, pages 620–629. IEEE Computer Society, 2005.

[38] Bugra Gedik and Ling Liu. Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *IEEE Trans. Mob. Comput.*, 2008.

[39] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. In *Proceedings of the 41st annual ACM Symposium on Theory of Computing (STOC)*, pages 351–360. ACM, 2009.

[40] Philippe Golle and Kurt Partridge. On the anonymity of home/work location pairs. In Hideyuki Tokuda, Michael Beigl, Adrian Friday, A. J. Bernheim Brush, and Yoshito Tobe, editors, *Proceedings of the 7th International Conference on Pervasive Computing (Pervasive 2009)*, volume 5538 of *Lecture Notes in Computer Science*, pages 390–397. Springer-Verlag, Nara, Japan, May 2009.

[41] Marco Gruteser and Dirk Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proc. of the First International Conference on Mobile Systems, Applications, and Services (MobiSys)*. USENIX, 2003.

[42] Marco Gruteser and Baik Hoh. On the anonymity of periodic location samples. In Dieter Hutter and Markus Ullmann, editors, *Proceedings of the Second International Conference on Security in Pervasive Computing (SPC 2005)*, volume 3450 of *Lecture Notes in Computer Science*, pages 179–192. Springer, 2005.

[43] Shen-Shyang Ho and Shuhua Ruan. Differential privacy for location pattern mining. In *Proceedings of the 4th ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS (SPRINGL)*, pages 17–24. ACM, 2011.

[44] Bret Hull, Vladimir Bychkovsky, Yang Zhang, Kevin Chen, Michel Goraczko, Allen Miu, Eugene Shih, Hari Balakrishnan, and Samuel Madden. Cartel: A distributed mobile sensor computing system. In *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems*, SenSys '06, pages 125–138. ACM, 2006.

[45] Mobile data challenge dataset. `https://www.idiap.ch/dataset/mdc`.

[46] Sibren Isaacman, Richard Becker, Ramón Cáceres, Margaret Martonosi, James Rowland, Alexander Varshavsky, and Walter Willinger. Human mobility modeling at metropolitan scales. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, MobiSys '12, pages 239–252. ACM, 2012.

[47] Oliver Jan, Alan Horowitz, and Zhong-Ren Peng. Using global positioning system data to understand variations in path choice. *Transportation Research Record: Journal of the Transportation Research Board*, 1725:37–44, 2000.

[48] John Krumm. Inference attacks on location tracks. In Anthony LaMarca, Marc Langheinrich, and Khai N. Truong, editors, *Proceedings of the 5th International Conference on Pervasive Computing (Pervasive 2007)*, volume 4480 of *Lecture Notes in Computer Science*, pages 127–143. Springer, 2007.

[49] John Krumm. A survey of computational location privacy. *Personal and Ubiquitous Computing*, 13(6):391–399, 2009.

[50] Location guard. https://github.com/chatziko/location-guard.

[51] Changsha Ma and Chang Wen Chen. Nearby friend discovery with geo-indistinguishability to stalkers. *Procedia Computer Science*, 34:352 – 359, 2014.

[52] Chris Y. T. Ma, David K. Y. Yau, Nung Kwan Yip, and Nageswara S. V. Rao. Privacy vulnerability of published anonymous mobility traces. In *Proceedings of the 16th Annual International Conference on Mobile Computing and Networking, (MOBICOM 2010)*, pages 185–196, 2010.

[53] Ashwin Machanavajjhala, Daniel Kifer, John M. Abowd, Johannes Gehrke, and Lars Vilhuber. Privacy: Theory meets practice on the map. In Gustavo Alonso, José A. Blakeley, and Arbee L. P. Chen, editors, *Proceedings of the 24th International Conference on Data Engineering, ICDE 2008, April 7-12, 2008, Cancún, México*, pages 277–286. IEEE, 2008.

[54] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), 2007.

[55] Microsoft Research (2012) GeoLife trajectories (v. 1.3) . `https://www.microsoft.com/en-us/download/details.aspx?id=52367`.

[56] Microsoft Trustworthy Computing. Location Based Services Usage and Perceptions Survey, January 2011. `https://www.microsoft.com/en-us/download/details.aspx?id=3250`.

[57] Darakhshan J. Mir, Sibren Isaacman, Ramón Cáceres, Margaret Martonosi, and Rebecca N. Wright. DP-WHERE: differentially private modeling of human mobility. In Xiaohua Hu, Tsau Young Lin, Vijay V. Raghavan, Benjamin W. Wah, Ricardo A. Baeza-Yates, Geoffrey C. Fox, Cyrus Shahabi, Matthew Smith, Qiang Yang, Rayid Ghani, Wei Fan, Ronny Lempel, and Raghunath Nambiar, editors, *Proceedings of the 2013 IEEE International Conference on Big Data, 6-9 October 2013, Santa Clara, CA, USA*, pages 580–588. IEEE, 2013.

[58] Mohamed F. Mokbel, Chi-Yin Chow, and Walid G. Aref. The new casper: Query processing for location services without compromising privacy. In Umeshwar Dayal, Kyu-Young Whang, David B. Lomet, Gustavo Alonso, Guy M. Lohman, Martin L. Kersten, Sang Kyun Cha, and Young-Kuk Kim, editors, *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB)*, pages 763–774. ACM, 2006.

[59] Mirco Musolesi, Mattia Piraccini, Kristof Fodor, Antonio Corradi, and Andrew T. Campbell. CRAWDAD data set dartmouth/cenceme (v. 2008-08-13). `http://crawdad.cs.dartmouth.edu/dartmouth/cenceme`, 2008.

[60] Kevin Orland. Stalker Victims Should Check For GPS. The Associated Press, February 2003. `http://www.cbsnews.com/news/stalker-victims-should-check-for-gps/`.

[61] Simon Oya, Carmela Troncoso, and Fernando Pérez-González. Back to the drawing board: Revisiting the design of optimal location privacy-preserving mechanisms. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, pages 1959–1972. ACM, 2017.

[62] Balaji Palanisamy and Ling Liu. Mobimix: Protecting location privacy with mix-zones over road networks. In Serge Abiteboul, Klemens Böhm, Christoph Koch 0001, and Kian-Lee Tan, editors, *Proceedings of the 27th International Conference on Data Engineering, ICDE 2011, April 11-16, 2011, Hannover, Germany*, pages 494–505. IEEE Computer Society, 2011.

[63] Pew research center: Internet & technology – mobile fact sheet, January 2017. `http://www.pewinternet.org/fact-sheet/mobile/`.

[64] Michal Piorkowski, Natasa Sarafijanovic-Djukic, and Matthias Grossglauser. CRAWDAD data set epfl/mobility (v. 2009-02-24). `http://crawdad.cs.dartmouth.edu/epfl/mobility`.

[65] Please Rob Me. `http://pleaserobme.com/`.

[66] Layla Pournajaf, Li Xiong, Vaidy Sunderam, and Xiaofeng Xu. Stac: Spatial task assignment for crowd sensing with cloaked participant locations. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '15, pages 90:1–90:4. ACM, 2015.

[67] QGIS Processing provider plugin. `https://github.com/SpatialVision/differential_privacy`.

[68] Luca Rossi, James Walker, and Mirco Musolesi. Spatio-temporal techniques for user identification by means of GPS mobility data. *EPJ Data Science*, 4(11), 2015.

[69] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, Nov 2000.

[70] Pierangela Samarati. Protecting respondents' identities in microdata release. *IEEE Trans. Knowl. Data Eng*, 13(6):1010–1027, 2001.

[71] Pierangela Samarati and Latanya Sweeney. Generalizing data to provide anonymity when disclosing information (abstract). In ACM, editor, *PODS '98. Proceedings of the ACM SIGACT–SIGMOD–SIGART Symposium on Principles of Database Systems, June 1–3, 1998, Seattle, Washington*, pages 188–188. ACM Press, 1998.

[72] Krishna Sampigethaya, Mingyan Li, Leping Huang, and Radha Poovendran. AMOEBA: Robust location privacy scheme for VANET. *IEEE Journal on Selected Areas in Communications*, 25(8):1569–1589, 2007.

[73] Zachary M. Seward. Tinder's privacy breach lasted much longer than the company claimed. Quartz Media LLC, July 2013. `https://qz.com/107739/tinders-privacy-breach-lasted-much-longer-than-the-company-claimed/`.

[74] Shashi Shekhar, Viswanath Gunturi, Michael R. Evans, and KwangSoo Yang. Spatial big-data challenges intersecting mobility and cloud computing. In *Proceedings of the Eleventh ACM International Workshop on Data Engineering for Wireless and Mobile Access*, MobiDE '12, pages 1–6. ACM, 2012.

[75] Reza Shokri. Privacy games: Optimal user-centric data obfuscation. *Proceedings on Privacy Enhancing Technologies*, 2015(2):299–315, 2015.

[76] Reza Shokri, Julien Freudiger, Murtuza Jadliwala, and Jean-Pierre Hubaux. A distortion-based metric for location privacy. In *Proceedings of the 8th ACM Workshop on Privacy in the Electronic Society (WPES)*, WPES '09, pages 21–30. ACM, 2009.

[77] Reza Shokri, George Theodorakopoulos, Jean-Yves Le Boudec, and Jean-Pierre Hubaux. Quantifying location privacy. In *IEEE Symposium on Security and Privacy*, pages 247–262. IEEE Computer Society, 2011.

[78] Reza Shokri, George Theodorakopoulos, George Danezis, Jean-Pierre Hubaux, and Jean-Yves Le Boudec. Quantifying location privacy: The case of sporadic location exposure. In *Proceedings of the 11th International Privacy Enhancing Technologies Symposium (PETS 2011)*, volume 6794 of *Lecture Notes in Computer Science*, pages 57–76. Springer, 2011.

[79] Reza Shokri, George Theodorakopoulos, and Carmela Troncoso. Privacy games along location traces: A game-theoretic framework for optimizing location privacy. *ACM Transactions on Privacy and Security*, 19(4):11:1–11:31, 2017.

[80] Reza Shokri, George Theodorakopoulos, Carmela Troncoso, Jean-Pierre Hubaux, and Jean-Yves Le Boudec. Protecting location privacy: optimal strategy against localization attacks. In Ting Yu, George Danezis, and Virgil D. Gligor, editors, *Proceedings of the 19th ACM Conference on Computer and Communications Security (CCS 2012)*, pages 617–627. ACM, 2012.

[81] Reza Shokri, Carmela Troncoso, Claudia Díaz, Julien Freudiger, and Jean-Pierre Hubaux. Unraveling an old cloak: k-anonymity for location privacy. In Ehab Al-Shaer and Keith B. Frikken, editors, *Proceedings of the 2010 ACM Workshop on Privacy in the Electronic Society, WPES 2010, Chicago, Illinois, USA, October 4, 2010*, pages 115–118. ACM, 2010.

[82] John Simerman. FasTrak to courthouse. East Bay Times, 2007. http://www.eastbaytimes.com/2007/06/05/fastrak-to-courthouse/.

[83] Yi Song, Daniel Dahlmeier, and Stéphane Bressan. Not so unique in the crowd: a simple and effective algorithm for anonymizing location data. In Luo Si and Hui Yang, editors, *Proceeding of the 1st International Workshop on Privacy-Preserving IR: When Information Retrieval Meets Privacy and Security*, volume 1225 of *CEUR Workshop Proceedings*, pages 19–24. CEUR-WS.org, 2014.

[84] Latanya Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5):571–588, 2002.

[85] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5):557–570, 2002.

[86] Kar Way Tan, Yimin Lin, and Kyriakos Mouratidis. Spatial cloaking revisited: Distinguishing information leakage from anonymity. In *Proceedins of the 11th International Symposium on Advances in Spatial and Temporal Databases (SSTD 2009)*. Springer, 2009.

[87] Galini Tsoukaneri, George Theodorakopoulos, Hugh Leather, and Mahesh K. Marina. On the inference of user paths from anonymized mobility data. In *Proceedings of the IEEE European Symposium on Security and Privacy (EuroS&P 2016)*, pages 199–213. IEEE, 2016.

[88] Amit Kumar Tyagi and N. Sreenath. A comparative study on privacy preserving techniques for location based services. *British Journal of Mathematics & Computer Science*, 10(4):1–25, 2015.

[89] Urban Airship. https://www.urbanairship.com/.

[90] Ting Wang and Ling Liu. *From Data Privacy to Location Privacy*, pages 217–246. Springer, Boston, MA, 2009.

[91] Marius Wernke, Pavel Skvortsov, Frank Dürr, and Kurt Rothermel. A classification of location privacy attacks and approaches. *Personal and Ubiquitous Computing*, 18(1):163–175, 2014.

[92] Yonghui Xiao and Li Xiong. Protecting locations with differential privacy under temporal correlations. In Indrajit Ray, Ninghui Li, and Christopher Kruegel, editors, *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS 2015)*, pages 1298–1309. ACM, 2015.

[93] Toby Xu and Ying Cai. Feeling-based location privacy protection for location-based services. In *Proceedings of the 2009 ACM Conference on Computer and Communications Security (CCS 2009)*. ACM, 2009.

[94] Andy Yuan Xue, Rui Zhang, Yu Zheng, Xing Xie, Jin Huang, and Zhenghua Xu. Destination prediction by sub-trajectory synthesis and privacy protection against such prediction. In *29th IEEE International Conference on Data Engineering (ICDE)*, pages 254–265. IEEE, 2013.

[95] Mingqiang Xue, Panos Kalnis, and Hung Pung. Location diversity: Enhanced privacy protection in location based services. In *Proc. of the 4th International Symposium on Location and Context Awareness (LoCA)*, volume 5561 of *Lecture Notes in Computer Science*, pages 70–87. Springer, 2009.

[96] Hui Zang and Jean 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 (MobiCom 2011)*, pages 145–156. ACM, 2011.

[97] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 791–800. ACM, 2009.

[98] Ge Zhong and Urs Hengartner. A distributed k-anonymity protocol for location privacy. In *Proceedings of the Seventh Annual IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–10. IEEE Computer Society, March 2009.

[99] Kathryn Zickuhr. Pew research center: Internet & technology – location-based services, September 2013. `http://www.pewinternet.org/2013/09/12/location-based-services/`.