TRANSACTIONS ON DATA PRIVACY 16 (2023) 51–82

# A Survey on Privacy in Human Mobility

**Anna Monreale***, **Roberto Pellungrini***

*Department of Computer Science, University of Pisa, Pisa, Italy.

E-mail: anna.monreale@unipi.it, roberto.pellungrini@di.unipi.it

Received 7 April 2022; received in revised form 22 July 2022; accepted 30 July 2022

**Abstract.** In the last years we have witnessed a pervasive use of location-aware technologies such as vehicular GPS-enabled devices, RFID based tools, mobile phones, etc which generate collection and storing of a large amount of human mobility data. The powerful of this data has been recognized by both the scientific community and the industrial worlds. Human mobility data can be used for different scopes such as urban traffic management, urban planning, urban pollution estimation, etc. Unfortunately, data describing human mobility is sensitive, because people's whereabouts may allow re-identification of individuals in a de-identified database and the access to the places visited by individuals may enable the inference of sensitive information such as religious belief, sexual preferences, health conditions, and so on. The literature reports many approaches aimed at overcoming privacy issues in mobility data, thus in this survey we discuss the advancements on privacy-preserving mobility data publishing. We first describe the adversarial attack and privacy models typically taken into consideration for mobility data, then we present frameworks for the privacy risk assessment and finally, we discuss three main categories of privacy-preserving strategies: methods based on anonymization of mobility data, methods based on the differential privacy models and methods which protect privacy by exploiting generative models for synthetic trajectory generation.

## 1 Introduction

An ever increasing number of technologies nowadays, uses or somehow manages data about our movements. In our daily lives we generate an astounding quantity of human movement data, by interacting with devices and applications that have become commonplace in our everyday life. Whenever we interact with our mobile phone, use navigation apps, access a social network etc., spatio-temporal points are generated describing where we are and when we are moving. The mobility of millions of people is tracked every day and used in a plethora of different services and applications. Some of them include: location based advertisements, traffic analysis, transportation systems design, behavior profile, car navigation systems and many more. The availability of human mobility data is not fundamental only for developing services based on intelligent systems but can be useful for understanding different and important phenomena which may have a huge impact on our society. For example, the analysis of human mobility following a natural disaster can help in improving the disaster management, in defining an effective humanitarian relief and long-term societal reconstruction [97]. And more, the analysis of the human mobility during the COVID-19 pandemic enables the understanding of the relationship between population's mobility and the viral transmissibility. This is important because can help in monitoring the pandemic's evolution and in dynamically adapting policy interventions

[18]. The availability of large quantities of mobility data is also at the forefront of research, both in academia and in industry [80, 24, 62].

The paradigm shift towards a *knowledge society*, i.e., a society where decisions can be taken – by individuals or by business and policy makers – on the basis of the knowledge distilled from the ubiquitous digital traces, comes with unprecedented *opportunities* and *ethico-legal risks*. Among these risks, *individual privacy* is one of the most important because it touches a human fundamental right.

Data is leaking everywhere, everyday: malicious entities can have access to millions of terabytes of data from a variety of sources, and the situation is getting worse. In 2021 alone there were more than 1,200 recorded data breaches [43] and the data of millions of people was exposed. For example, a breach in the popular social network Linkedin exposed the details of more than 700 million profiles. The damage of this data breaches is twofold: not only the information is now leaked and in the open, with the privacy of many individuals irremediably damaged, but now this information can also be used as a knowledge pool for future, possibly even worse attacks. As a consequence, privacy is one of the top concerns for both customers and professionals across multiple fields, especially data management [89, 19]. Public opinion's awareness of privacy issues in data management and analysis processes is constantly growing and people are getting educated more on the matter each day. People's awareness is particularly raising for privacy risks derived from the use and processing of mobility and location data [109, 94]. Unfortunately, this skepticism is well founded because the particular nature of mobility data may derive the inference of very sensitive information on the personal and private sphere. Locations visited by an individual in combination with the time of the visits can reveal not only personal habits but also personal preferences related to sensitive aspects such as religion, health status, etc. Pseudonymization of mobility traces, obtained by hiding direct identifiers simply replacing them with pseudonyms, is not a solution because some mobility patterns hidden in the data itself can reveal unique behaviors enabling individual re-identification [21].

However, today the paradoxical situation we are facing is that we are fully running the risks, without fully catching the opportunities of our personal (mobility) data: while we feel that our private personal sphere is vanishing in the digital world, and that our personal data can be used without feedback and control, the same data are seized in the databases of companies (e.g., telecommunication companies, insurance companies, etc), which use legal constraints on privacy as a reason for not sharing it with science and society at large. This leads to an absurd situation where we cannot benefit from our data and this precious source of knowledge is locked to data analysts or service developers. What happened, for example, with the data regarding COVID-19 contact tracing is emblematic: there has been a huge difference in policy between countries concerning the use of the data. Many experts advocate for implementation of innovative, privacy preserving techniques to allow the meaningful use of this data while at the same time protecting the privacy of the individuals represented. [74].

In Europe, policy-makers reacted to this important issue by updating the European legislation, replacing the 1995 Data Protection Directive with the General Data Protection Regulation (GDPR)[82]. Such regulation addresses privacy threats raised by the processing of human and personal data by strengthening protections for individuals requiring the application of the *privacy-by-design* principle, and by explicitly recognizing location data as a factor enabling individual re-identification (Article 4 of GDPR). Moreover, since many sensitive attributes are uniquely associated with places, collecting and publishing mobility data that show a person frequently visits a place or attends a particular event represents a means to draw a comprehensive picture of an individual and to lead to classify in that case

*location data* as a special category of personal data or *sensitive data* requiring high level of protection (Article 9 of GDPR). A very interesting study on the location data privacy under the GDPR can be found in [73].

For guaranteeing the applicability of this regulation, we need to enable knowledge discovery from data by setting mobility and location data *free*, i.e., it becomes fundamental to have tools for exploiting the advantage of analyzing this type of data while *by-design* preventing privacy violations, which may result in negative economic and social impacts.

**Contribution and survey organization.** Research activity on privacy issues and mitigation strategies for mobility data has been very active in the last two decades and today it is still a prolific field. This confirms the great interest of the scientific research community and industrial world on the power associate to the availability of mobility data. The research literature on privacy for mobility data is covered by two main research lines: privacy in location based services and privacy-preserving trajectory data publishing.

The first research line addresses the problem of protecting the user location privacy in real-time while answering a query generated by a mobile device [9], while the second one studies the problem of making available mobility datasets while protecting privacy of individuals represented in the data [10].

In this survey paper we give an overview of the main results in the second research area. Designing privacy-preserving strategies for trajectory data requires to manage the trade-off between privacy protection and data utility, an objective that in this context is particularly challenging because of the high dimensionality of these type of data. Although the data utility is hard to maintain under control it is very important because private mobility data after the sanitization process should be used as a proxy of human behavior which may enable the understanding of important social and complex phenomena. A privacy-preserving technique that completely destroys the human mobility laws hidden in the data would make them useless.

This survey discusses three main aspects related to the privacy in mobility data: *(i)* the privacy attack models, that a malicious adversary may conduct on trajectory data, and the privacy models that guide the privacy mitigation strategies; *(ii)* the privacy risk assessment frameworks designed for simulating the privacy attacks and quantifying the privacy risks inherent to a mobility dataset under analysis; and *(iii)* the mitigation strategies developed to counter the well-known attacks. Concerning the privacy-preserving strategies we will overview: methods assuring privacy by $k$-anonymity based models and its variants, methods based on the differential privacy model and methods exploiting deep leaning generative models for protecting privacy. We also discuss recent privacy-preserving methods that address the challenging task to preserve privacy in a setting where each individual can express a personal privacy requirement (expectation). Although in the literature some other survey exist [32, 50], to the best of our knowledge, this is the first survey that includes in the literature overview also methods for personalized privacy in mobility data, methods that protect individual privacy by generating synthetic trajectories by deep learning models and frameworks for privacy risk assessment.

The remaining of the paper is organized as follows. Section 2 introduces some preliminary notions and definitions related to mobility data. Section 3 overviews the adversarial attacks for mobility data and the privacy models exploited by the privacy-preserving techniques. In Section 4 we describe the most important frameworks for a quantitative assessment of privacy risks while in Section 5 we discuss the state-of-the-art privacy protection strategies for trajectory data. Lastly, Section 6 concludes this survey.

# 2 Mobility data

The techniques presented in this survey all focus on human mobility data, i.e., data describing the movements of a set of individuals during a period of observation. This type of data is generally collected in an automatic way through electronic devices (e.g., mobile phones, GPS devices) in form of raw trajectory data. A raw trajectory of an individual is a sequence of records identifying the movements of that individual during the period of observation [121]. A trajectory can therefore be represented as the list of spatio-temporal points describing the movement of the individual.

**Definition 1 (Trajectory).** The trajectory $T_u$ of an individual $u$ is a temporally ordered sequence of tuples $T_u = \langle (l_1, t_1), (l_2, t_2), \ldots, (l_n, t_n) \rangle$, where $l_i = (x_i, y_i)$ is a location, $x_i$ and $y_i$ are the coordinates of the geographic location, and $t_i$ is the corresponding temporal information, $t_i < t_j$ if $i < j$. Locations are generally ordered according to the temporal information, from the least recent to the most recent in time.

Intuitively, each pair $(l_i, t_i)$ indicates that the moving object is in the position $l_i = (x_i, y_i)$ at the specific time $t_i$.

**Definition 2 (Sub-Trajectory).** Let $T_u = \langle (l_1, t_1), (l_2, t_2), \ldots, (l_n, t_n) \rangle$ be a trajectory. A trajectory $S = \langle (l'_1, t'_1), (l'_2, t'_2), \ldots, (l'_m, t'_m) \rangle$ is a sub-trajectory of $T_u$ or is contained in $T_u$ ($S \preceq T_u$) if there exist integers $1 \leq i_1 < \ldots < i_m \leq n$ such that $\forall 1 \leq j \leq m (l'_j, t'_j) = (l_{i_j}, t_{i_j})$.

Many works proposing human mobility analysis or privacy-preserving techniques for trajectory data do not exploit in depth the temporal information, they only consider a trajectory as a ordered sequence of locations, i.e., $T_u = \langle l_1, l_2, \ldots, l_n \rangle$.

A *Mobility Dataset*, denoted by $D$, is a set of trajectories $\{T_1, \ldots, T_N\}$ referred to $N$ individuals or moving objects.

| ID | Trajectory | Diagnosis |
|---|---|---|
| $u_1$ | $\langle l_a, l_b, l_c \rangle$ | Flue |
| $u_2$ | $\langle l_e, l_f, l_a, l_c \rangle$ | AIDS |
| $u_3$ | $\langle l_a, l_b, l_a, l_f, l_a \rangle$ | Diabetes |
| $u_3$ | $\langle l_a, l_b \rangle$ | AIDS |

Table 1: Trajectory data and health data

Sometimes trajectory data are published together with some other sensitive attributes, which not have a spatio-temporal nature, for instance attributes describing diseases, religion belief, etc. Table 1 shows an example of these trajectories. Alternatively, we can also have trajectory data where geographical positions are enriched with semantic information that describe aspects like the reason of the visit of a position/region. Typically, given a raw trajectory some stops are identified and enriched with semantic information (e.g., hotels, restaurants, museums, etc.) derived from a predefined taxonomy that is application dependent. The derived trajectory is called *semantic trajectory* [72]. Figure 1 illustrates the concept of semantic trajectory corresponding to a raw trajectory. In the semantic trajectory the moving object first was at *Home* (stop 1), then she went to *Hospital* (stop 2), later she went to *Work* (stop 3), and finally the moving object went to the *Gym* (stop 4).

This type of trajectories can reveal very sensitive information associated to the movement. For example, the trajectory in Figure 1 reveals that the individual visited the *Hospital* and this could help an attacker to infer sensitive information about diseases. Some works in the literature studied privacy issues also for this particular type of trajectory data [72].

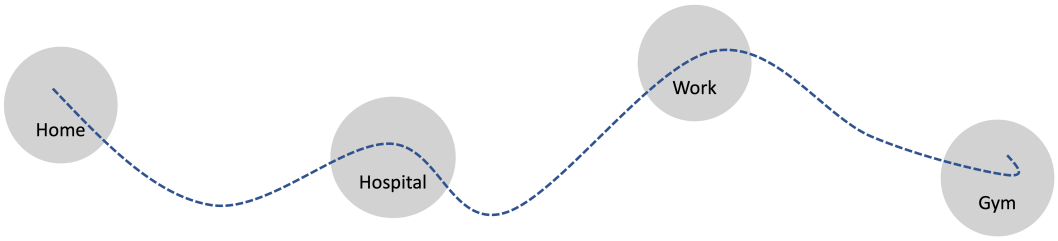


Figure 1: Example of Semantic Trajectory

Table 2: Summary of privacy attacks on mobility data.

| Authors & Reference | Year | Attack Model | Adversary Knowledge |
|---|---|---|---|
| Krumm [54] | 2007 | Record Linkage | Home location |
| Abul et al. [1] | 2008 | Record Linkage | Sub-trajectory & spatial-time approx. |
| Mohammed et al. [69] | 2009 | Record and Attribute Linkage | Specific sub-sequences |
| Yarovoy et al. [114] | 2009 | Attribute Linkage | Sub-trajectory |
| Monreale et al. [70] | 2010 | Record Linkage | Sub-trajectory |
| Zang and Bolot [117] | 2011 | Record Linkage | Top-$n$ frequent locations |
| Freudiger et al. [34] | 2011 | Attribute Linkage | Sub-trajectory data |
| Monreale et al. [72] | 2011 | Attribute Linkage | Taxonomy of visits |
| Srivatsa and Hicks [98] | 2012 | Record Linkage | Side social-graph |
| de Montjoye et al. [21] | 2013 | Record Linkage | Sampled spatio-temporal points |
| Chen et al. [15] | 2013 | Attribute Linkage | Sub-trajectory |
| Rossi and Musolesi [92] | 2014 | Distance Based Record Linkage | Location-based data |
| Chen et al. [16] | 2015 | Probabilistic Attack | Mobility profiles |
| Sui et al. [100] | 2016 | Attribute Linkage | Sub-trajectory |
| Gramaglia et al. [38] | 2017 | Probabilistic Attack | Sub-trajectory |
| Liu et al. [60] | 2018 | Attribute Linkage | Sub-trajectory with a max length |
| Yao et al. [113] | 2019 | Similarity Attack | Sub-sequence data with a max length |
| Tu et al. [108] | 2019 | Semantic Attack | PoI distribution |
| Basik et al. [8] | 2020 | Distance Based Record Linkage | History of movements graphs |

# 3 Adversary Attacks and Privacy Models

In the context of mobility data, a privacy attack is usually defined as the attempt by a malicious adversary to infer, understand or somehow extrapolate information about an individual's mobility by leveraging some knowledge about that individual. The way in which this knowledge is built and then leveraged varies depending on how the adversary is modeled and on how the actual attack is modeled. There are three main families of attacks. In this section we summarize the literature regarding them.

## 3.1 Record Linkage Attacks

The attacks belonging to this category are based on the same basic structure: an adversary, i.e. a malicious entity intent on breaching the privacy of an individual, gets access to some knowledge regarding the target individual. This knowledge can be generally assumed to be some portion of data relating to the individual, and can be gathered by an adversary in a variety of ways, either through direct observation of the individual or by the collection of separate data sources. This is called "prior knowledge" or "background knowledge" [35] and the corresponding kind of attack is also referred to as **Background Knowledge Attack**.

The adversary uses the knowledge gathered to perform an attack on some anonymized available dataset. The goal of the adversary is to correctly match the data in her possession to the record in the anonymized data corresponding to the actual individual of the background knowledge. If the adversary successes in her scope, she can re-identify the individual in the data and/or infer new information about the individual thus breaching the individual's privacy. This may be a particularly damaging outcome in the specific case of mobility data, as the information contained in a trajectory is very sensitive: knowing where and when an individual moves can allow multiple inferences regarding personal details, habits, behaviors etc.

Terrovitis et al. [103] and Xu et al. [112] introduced the "prior knowledge based" attacks on transactional data. The general concept was next applied in the context of mobility data in several works taking into consideration the sequential nature of this type of data: for example, Mohammed et al. [69] introduced a background knowledge attack where the adversary knows the exact sequence of the visited locations as they appear in the trajectory of the target individual, an attack later used by Monreale et al. [70]. Another example is the attack introduced by Abul et al. in which the adversary also knows information regarding the time of visit of an individual [1].

In these type of attacks, the assumption is that the knowledge of the adversary is itself present inside the data under attack. In essence, this means that the background knowledge can be simulated with a sub-sampling strategy applied to the trajectory data under analysis. Therefore, following our Definitions 1 & 2, the background knowledge of an adversary $\mathcal{B}_u$ can be described as the sub-trajectory of a given individual $u$, i.e., $\mathcal{B}_u \preceq T_u$. One notable example of simulation of a such kind of attack is presented by de Montjoye et al. in [21], where human mobility trajectories are attacked by subsampling random points from the trajectories themselves. A special case of sub-sampling strategy is to extract particular locations visited by an individual and use them for attacking the anonymous data. For instance, Krumm [54] uses the knowledge about the inferred *home* of the individual target of the attack to perform the trajectory linkage while Zang et al. [117] exploit the top-$n$ frequent locations visited by an individual to conduct the attack. Here it is therefore assumed that the adversary has knowledge regarding the frequency of visit.

Other record linkage attacks aim at cross-referencing information from data related to the same individual represented in the trajectory but stored in other databases. For example, Srivatsa and Hicks [98] leverage social graphs to extract background knowledge to perform an attack on individuals in a mobility dataset, considering that individuals that are friends in the social network may encounter each other more frequently in the mobility dataset.

In case the information to be used for as background knowledge is derived from another database, a powerful type of record linkage attack is the **Distance Based Record Linkage Attack**[106]. The basic principle of this attack is that an adversary, having access to some data, tries to link her data to specific records in an available database applying a distance function to the two data sources. This technique is commonly used in database integration, when multiple data sources belonging to the same entities need to be consolidated into a single comprehensive source. The distance function is the central point of this approach: the most popular distance function used is the Euclidean one, but many other functions have been tested. As an example, Rossi and Musolesi in [92] performed an analysis on distance based linkage of trajectory data from location-based social networks, applying the Hausdorff distance between the locations in the trajectories. Another recent development on this particular kind of attack is the *SLIM* algorithm by Basik et al. [8]. The SLIM algorithm works on the history of movements of each individual and represents similarities between histories of movements as graphs, weighting each connection with a defined sim-

ilarity (i.e. inverse of the maximum geographical distance over a given time-frame).

## 3.2 Attribute Linkage Attacks

Attribute Linkage attacks are also known as **homogeneity attacks** [35]. Here, the adversary tries to link the background knowledge to some sensitive attributes in the data. In particular, the goal is not to re-identify the specific individual record but infer sensitive information within the record. As a consequence, the non-unicity of the records with respect the quasi-identifiers, used for linking the adversary background knowledge, could be not enough to counter the attack in case we have a lack of diversity for the sensitive information. As an example, if in a tabular context we have 100 records with the same value of quasi-identifiers gender, city, age the inference of the individual disease is unfortunately enabled by the sharing of the same value of the disease.

In the field of trajectory data the homogeneity attack can be applied to infer sensitive locations [100, 34] or sensitive attributes explicitly or implicitly associated to each trajectory [69, 15]. In the first case, it is assumed that some locations are considered more sensitive with respect to others and the sensitivity can be derived by the semantic associated to the location. Indeed, if we consider the possibility to infer points of interest (PoIs) and the scopes of the different stops in a movement (see the notion of Semantic Trajectory in Section 2) we can easily understand that some stops could lead to more sensitive inference with respect to others. For example, visiting an *hospital* may lead to more sensitive inference than visiting a *park*. Some works assume that the sensitivity of some locations can be defined by the individuals [114] or by domain experts [72]. Other works [100, 34] show that the semantic of some locations can be easily inferred by analyzing the accessed trajectories by exploiting the information about regular and frequent visits of some locations. This is particularly harmful, since often the most frequent locations correspond to the *home* and *work* locations of an individual. In order to guarantee the diversity of the sensitive attributes or sensitive locations some works limit the length of the locations to be used as background knowledge [69, 61].

## 3.3 Probabilistic Attacks

Another family of attacks that can be applied to mobility data are probabilistic attacks, also known as **inference attacks** in the literature. In this kind of attack, a malicious adversary tries to increase her knowledge by accessing the target database [35]. As a consequence, to guarantee protection against this attack, the access to a mobility dataset should not reveal too much additional information with respect to what is already known by the adversary. This type of attack can be seen as a generalization of the attribute linkage attack because it measures the increase of knowledge, not only in terms of disclosure of sensitive attributes and locations, but in terms of quantity of unknown information acquired. An example of probabilistic attack on trajectory data has been defined in the work of Gramaglia et al. [38], where the adversary success is measured in terms of the number of additional locations acquired after the attack. Also the **similarity attack** [113] belongs to this category. It indeed exploits the similarity between different sensitive attribute values to infer sensitive information also in presence of diversity of the values. For example, if we have three individuals with the same trajectory data and with associated three different diseases *gastric ulcer*, *gastritis* and *stomach cancer* an adversary may conclude that the individual under attack has some *stomach-related problems*, given that all the diseases belong to the same category.

Similarly, the **semantic attack** enables an increase of the adversary knowledge exploiting the semantic information associated to some spatio-temporal points of the trajectories. For example, one could derive frequent visits of an individual to a region where most of the POIs are related to public or private health institutions as a consequence the analysis of the individual mobility together with the POIs distribution can cause a privacy disclosure revealing information about the individual health status [108].

The **activity attack** proposed by Chen et al. [16] follows a similar principle: here, the adversary uses the mobility profiles of individual to estimate the probability of an activity trajectory, i.e., a sequence of visits with a certain staying time, for a number of PoIs for each individual. Here the focus is not on the actual spatio-temporal information, but on what the individual did in its trajectory, that is, the purpose of the movement. The authors use an optimization approach to find the most likely activity trajectory for ech individual.

## 3.4 Privacy Models

In this section we introduce the privacy models and their standard definitions that are used in many techniques addressing the problem of guaranteeing privacy in trajectory data against the attacks described in the previous section. We highlight that in some cases, for instantiating these models in the field of mobility data, it has been required to define opportune variants suitable to the needs of the particular nature of trajectory data. These variants will be discussed in Section 5.

**$k$-anonymity.** To counter record linkage attacks most of the privacy-preserving techniques base their protection strategy on the $k$-anonymity privacy model [93] or its variants that often are necessary to guarantee higher data quality. This model has been initially introduced for tabular data [93] and later has been applied to different types of data such as mobility data [1, 107, 70, 71], itemset-based data [103, 6], sequence data [71] and graph data [99].

The standard $k$-anonymity privacy model assumes that the set of attributes in a dataset is divided into *sensitive* attributes and *quasi-identifiers*. Sensitive attributes are the attributes that need to be protected. Quasi-identifiers are attributes that may be linked to external information retrieved by an adversary for a linking attack. If the adversary succeeds in the linking then, can get access to the identity of the individual and its sensitive attributes. Therefore, this privacy model requires that for each released record we have at least $(k-1)$ other records in the released dataset whose values are indistinct over the quasi-identifiers.

The $k$-anonymity is a property that can be achieved on a given dataset using techniques based on: *generalization* [46], i.e. reducing the granularity of the representation of quasi-identifiers; *suppression* [101], i.e. deleting the value of highly informative attributes from the data altogether, and *microaggregation*[25], i.e., a perturbation-based protection method where the data is divided into small clusters and values of quasi-identifier attributes are substituted with the values of the centroids of the clusters. The problem of achieving optimal $k$-anonymity has been proven to be NP-Hard in [67]. Some heuristics have been proposed in the literature to achieve $k$-anonymity. For example, a greedy partition-based algorithm was proposed in [56].

**$l$-diversity.** $K$-anonymity has some vulnerabilities: it does not provide protection against attribute linking attacks. To tackle this problem, Machanavajjhala et al. in [63] propose the $l$-diversity model, with the objective of maintaining a degree of diversity in the sensitive

attributes of an anonymity set, i.e., a group pf $k$ record sharing the same values of quasi-identifiers. The $l$-diversity principle requires that every group of individuals, that can be isolated by an attacker by a specific background knowledge exploiting the quasi-identifiers, should contain at least $l$ well-represented values for a sensitive attribute. A number of different instantiations for the $l$-diversity definition have been also proposed for complex data such as social network data [122] and mobility data [69, 15, 72].

**$t$-closeness.** However, in case the overall distribution of the sensitive attribute is skewed, further measures have to be taken to prevent inference of sensitive information. In this case the *t-closeness* model, introduced in [59], could help in making data safe against these attacks. This privacy model imposes that the distance between the distribution of a sensitive attribute in any equivalence class and the distribution of that attribute in the overall dataset has to be bounded by a threshold $t$. Also this model, as we will see in Section 5.1.2, has been successful applied to trajectory data to counter the semantic attack [108].

**Differential Privacy.** In 2006, Dwork et al. [27] introduce the Differential Privacy model that can be satisfied applying specific techniques of data randomization. The fundamental idea at the base of differential privacy is that an algorithm applied to two datasets that differ only on the record of a single individual should yield almost the same result. This means that the individual can safely submit her record to the dataset because nothing, or almost nothing, can be discovered from the database with her information that could not have been discovered without her information. More formally, a randomized algorithm $A$ is $\epsilon$-differentially private if for all datasets $D$ and $D'$ differing only on a single record, and for all $S \subseteq range(A)$ the property $Pr[A(D) \in S] \leq e^{\epsilon} \times Pr[A(D') \in S]$ holds. A relaxed version of differential privacy was proposed in [7], in which the authors claim that the privacy protection can be achieved even when admitting a small amount of privacy loss. Formally, the relaxed version of the differential privacy model, named $(\epsilon, \delta)$-differential privacy, changes then to $Pr[A(D) \in S] \leq e^{\epsilon} \times Pr[A(D') \in S] + \delta$. Note that, $\delta$ models the privacy loss and with $\delta = 0$ we have the original definition of differential privacy. The two most common mechanisms for achieving different privacy are the *Laplace* and *Exponential* mechanisms. The first one requires to add noise to the results of the algorithm to be computed drawing it from a Laplace distribution [28]. The noise has to be proportional to the global sensitivity of the algorithm under consideration. This approach is suitable for randomizing numerical values. There are however cases in which adding noise through the Laplace distribution may not be feasible. For the analysis whose outputs are not real or when adding noise destroys the sense of data, the authors of [66] propose an exponential mechanism, selecting an output from the output domain, $r \in R$, by taking into consideration the score of a given utility function $q$. In [29] authors give estimates for the $\epsilon$ parameter, stating how it is possible to produce meaningful results even assuming values larger than 1. This privacy model has been extensively used in different contexts. In Section 5.2 we overview the literature that applies this model and its variants to trajectory data especially to counter probabilistic attacks.

# 4 Privacy Risk Assessment

**Privacy risk assessment**, also called **Privacy impact assessment** is a crucial part in any Privacy By Design process [12, 11]. Privacy risk assessment is traditionally defined in

generic terms, with little reference to practical methodologies. Some examples are the NIST methodology [47] or the OWASP [79] risk rating methodology. These methodologies are often cited as guidelines, even though they lack substantial indications on the actual techniques to be used and how to tackle the specific nature of certain kinds of data. Privacy threats are hereby ranked depending on their perceived likelihood or potential damage. What we are interested in is methodologies to quantify privacy risk with some metric. Wagner compiled a comprehensive survey on the metrics used to quantify privacy risk [111] and the same Wagner with Eckhoff performed several simulations of privacy assessment based on attack models for vehicular movement data [110]. Among these, for example, the User Centric Privacy metric introduced by Freudiger et al. [33] measures privacy risk based on the time elapsed from the last application of a privacy protection countermeasure.

In general, we can assess privacy risk by evaluating how much some attack can be effective on a certain dataset. This means that, to properly understand if individuals in a mobility dataset are at risk, the most accurate way is to simulate some attacks on the data and evaluate how an adversary would fare. This way of assessing privacy risk is, however, heavily dependent on the context and chosen threat model. Shokri et al. model the adversary attack against location privacy preserving techniques by formalizing its performance with three different metrics: *accuracy*, *certainty*, and *correctness* [95]. They state that *correctness*, i.e., the inverse of the probability of the adversary error, is the metric that determines the privacy of users. This is something that is often expressed in many other works in various ways: an attack is more powerful when its probability of failure is minimized.

In privacy literature, the general assumption when simulating a privacy attack is the **worst-case scenario**, i.e., it is assumed that an adversary could use all possible knowledge in her attack. Pratesi et al. proposed PRUDEnce [88], a framework for assessing both the empirical (not theoretical) privacy risk associated to users represented in the data, and the data quality guaranteed only with users not at risk. The framework considers a scenario where a Data Analyst requests a Data Provider human mobility data in order to develop an analytical service. For its part, the Data Provider has to guarantee the right to privacy of the individuals whose data are recorded. As a first step, the Data Analyst communicates to the Data Provider the data requirements for the analytical service. Assuming that the Data Provider stores a database $\mathcal{D}$, it aggregates, selects and filters the dataset $\mathcal{D}$ to meet the requirements by the Data Analyst and produces a set of mobility datasets $\{D_1, \ldots, D_z\}$ each with a different data structure and/or aggregation of the data. The Data Provider then reiterates a four-step procedure until it considers the data delivery safe:

1. *Identification of Attacks*: identify a set of possible attacks that an adversary might conduct in order to re-identify the individuals in the mobility datasets $\{D_1, \ldots, D_z\}$;
2. *Privacy Risk Computation*: simulate the attacks and compute the set of privacy risk values for every individual in the mobility datasets $\{D_1, \ldots, D_z\}$;
3. *Dataset Selection*: select a mobility dataset $D \in \{D_1, \ldots, D_z\}$ with the best trade-off between the privacy risks of the individuals and the quality of the data, given a certain level of tolerated privacy risk and the data requirements by the Data Analyst;
4. *Risk Mitigation and Data delivery*: apply a privacy-preserving transformation (e.g., generalization, randomization, etc.) on the chosen mobility dataset $D$ to eliminate the residual privacy risk, producing a filtered mobility dataset $D_{filt}$. Deliver the mobility dataset $D_{filt}$ to the Data Analyst when the $D_{filt}$ is adequately safe.

The framework is summarized in Figure 2.

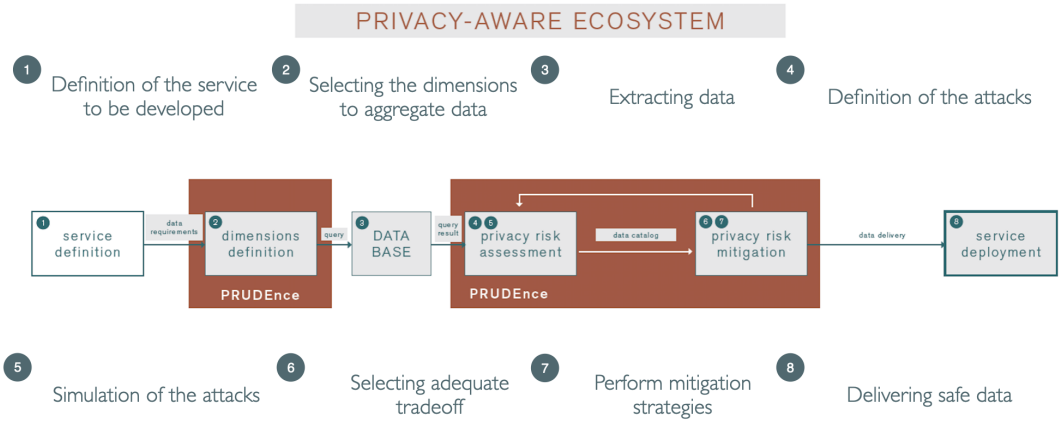


Figure 2: PRUDEnce privacy-aware ecosystem schema.

The framework allows for the definition of the background knowledge of an adversary in terms of the dimensions of the data available (in the context of mobility data this could be for example the time and/or the frequency of visits to a location). In PRUDEnce, attacks are evaluated based on the k-anonymity achieved by individuals in the data, i.e., how well hidden individuals are within each other with respect to the knowledge of the adversary. The simulation of an attack is done for every possible knowledge that the adversary may have, thus guaranteeing the **worst-case scenario** principle. In the same paper, Pratesi et al. show an application of the framework to mobility data. For how thorough PRUDEnce is, this comes at the cost of high computational complexity, being a combinatorial evaluation algorithm. Moreover, as new data gets added or previous data is updated, risk needs to be recomputed from the start. To overcome this issue Pellungrini et al. [83] proposed a data mining approach to predict privacy risk in mobility data based on individual mobility profiles. The basic idea is to use machine learning models such as Random Forests to predict privacy risk so that PRUDEnce is needed just as a one-time labeller for the data. The approach was tested on mobility profiles extracted from data covering mobility of two major cities in Tuscany for a period of observation of one-month. The results show that indeed, privacy risk may be predicted from individual mobility profile, and that the approach is transferable between regions with similar mobility patterns. Given the recent trends in explainable AI, Naretto et al. proposed an extension to the PRUDEnce framework called EXPERT [75, 76]. EXPERT (EXplainable Privacy ExposuRe predicTion), exploits machine learning models for predicting a user's individual privacy risk and local explainers for producing explanations of the predicted risk. First, EXPERT extracts from human mobility data an individual mobility profile describing the mobility behavior of any user. Second, for each user it exploits PRUDEnce to compute the associated privacy risk. Third, it uses the mobility profiles of the users with their associated privacy risks to train a machine learning model. For the prediction task, EXPERT exploits tree-based ensemble models to effectively handle the class-imbalance problem, i.e., a high number of risky users vs a low number of non-risky ones, that is typical of the data in this context. The aim is to have a predictor that preserves the privacy of risky users while providing the freedom of using data-driven services to users with low privacy risk. For a new user, along with the prediction of risk, EXPERT also provides an explanation of the predicted risk. EXPERT exploits two state-of-the-art explanation techniques, i.e., SHAP and LORE. The two methods pro-

duce explanations based on feature importance and logic rules, respectively. The goal of explanations is to provide users with insights on which mobility behavior contributes to their privacy risk. EXPERT is summarized in Figure 3

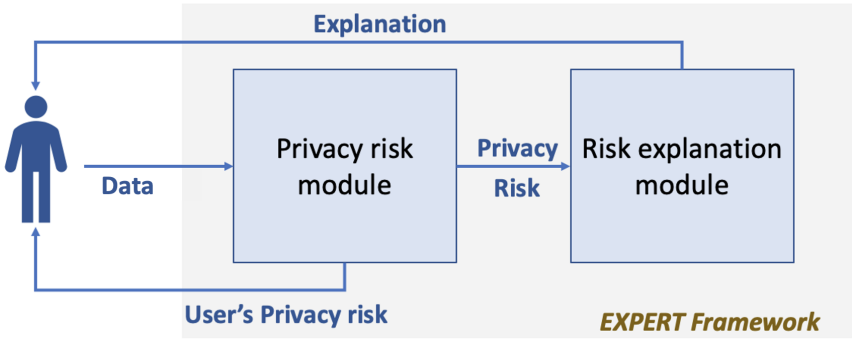


Figure 3: EXPERT Framework.

All the frameworks mentioned above assess privacy risk by simulating background knowledge based attacks. However, no assumption is made on the actual availability of the assumed knowledge for the adversary: it is systematically generated from the original data under attack. Pellungrini et al. [84] propose an adversarial model tailored for human mobility data, where the background knowledge is generated by modeling the behavior of an adversary as a mobility trajectory. The authors present three scenarios:

1. if the adversary is assumed to be one of the individuals in the data, the adversary trajectory is one of the original trajectory in the data;
2. if the adversary is assumed to be an individual not present in the data, the adversary trajectory is synthesized with a generative method trained on the original data;
3. if the adversary is assumed to be actively trying to maximize the overall knowledge in the data, the adversary trajectory is built with an optimization method. For this last scenario, the authors propose an adaptation of Simulated Annealing to mobility data, with the addition of spatio-temporal constraints. The end goal of the method is to generate a trajectory that, if followed by the adversary, would maximizes the knowledge of an adversary while at the same time maintaining plausible movements in the considered area.

The authors prove that, when constrained to plausible movements, it would be hard for the adversary to target multiple individuals in an area as it would require him to survey multiple locations and adopt peculiar patterns of movement. Moreover, the privacy risk for the individuals is considerably lower than the theoretical worst-case scenario.

Other privacy risk assessment frameworks have been proposed in the literature. ARX [87, 86] for example is an anonymization and privacy risk assessment tool, where the assessment is done to verify the efficacy of the privacy preserving techniques applied on the data. In essence, this means that the data is sanitized with some technique and metrics are applied to the data in order to verify the efficacy of this technique. Currently, ARX in mainly applied to tabular data and struggles with high dimensionality (30 attributes dataset is considered high-dimensional). PRUDEnce is instead designed to be applied alongside privacy protection techniques, to select the best approach to protect the data

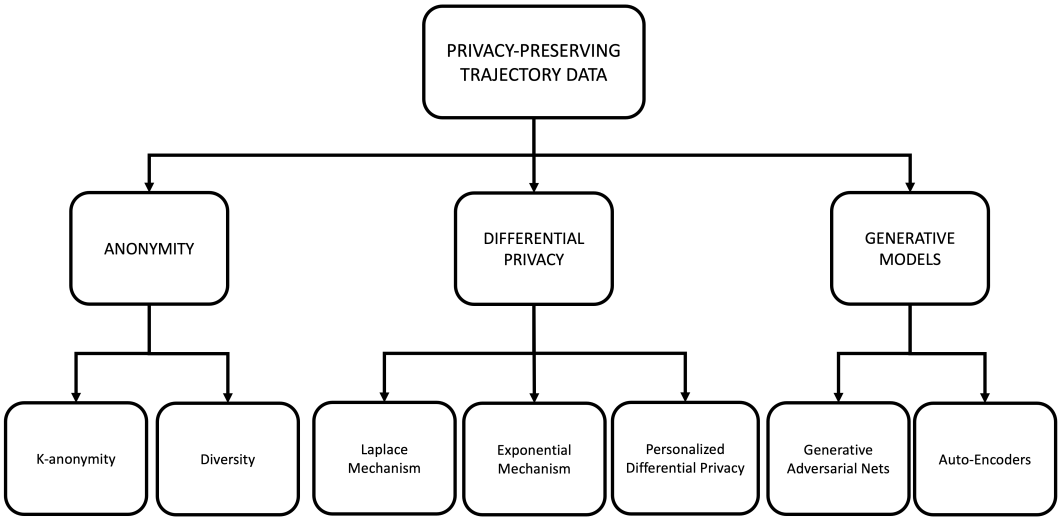


Figure 4: Taxonomy of privacy-preserving techniques for trajectory data

and maximize utility. Makri et al. [65] propose a privacy impact assessment methodology based on the characteristics of the organization responsible for the data. However, from a purely technical standpoint, this approach is a variation of the well known CRAMM [30] and therefore does not provide significant technical information to be readily applied.

# 5 Privacy-Preserving Methods

Concerning the privacy-preserving techniques for trajectory data, as summarized in Figure 4, the analysis and review of the literature led to the identification of three main categories: anonymity based techniques (summarized in Table 3), differential privacy based techniques (listed in Table 4) and methods exploiting generative models for protecting privacy (summarized in Table 5). All the methods described in the following aim at protecting individual privacy while publishing mobility data, which can be then freely used for other scopes, such as developing data-driven services and studying social and complex phenomena (e.g., social well-being [81], virus spread [4], human predictability [80, 24], estimation of air pollution [78], estimation of migration flows [96], etc.).

## 5.1 Privacy protection by Anonymity

Anonymity based protection techniques have the goal of reducing the probability of trajectory identification by public information. They are suitable in case trajectory data transformation for privacy protection is not to be performed at data-collection time. In the literature related to temporal data privacy, three privacy models have been taken into consideration: $k$-anonymity, $l$-diversity and $t$-closeness.

### 5.1.1 Approaches based on $k$-Anonymity and its variants

Among the privacy models, $k$-anonymity [93] is one of the most extensively applied in trajectory anonymity (see Section 3.4 for more details). The goal of $k$-anonymity is to guarantee that every individual object is hidden in a crowd of size $k$. In the standard tabular setting, a dataset satisfies the property of $k$-anonymity if each released record has at least ($k$-1) other records also visible in the release whose values are indistinct over the *quasi-identifiers*; i.e., attributes that, in combination, can uniquely identify individuals, such as birth date and gender. In $k$-anonymity techniques, methods such as *generalization* and *suppression* are usually employed to reduce the granularity of representation of *quasi-identifiers*. The method of generalization generalizes the attribute values to a range in order to reduce the granularity of representation. For instance, the city could be generalized to the region. Instead, the method of suppression, removes the value of an attribute.

In the case of trajectory data, where each record is a temporal sequence of location visited by a specific person, the above dichotomy of attributes into quasi-identifiers (QI) and private information (PI) does not hold any longer: here, a (sub)sequence of locations can play both the role of QI and the role of PI [71]. To see this point, consider the case where the attacker may know a sequence of locations visited by some specific person $P$: e.g., by shadowing $P$ for some time, the attacker may learn that $P$ was in the shopping mall, then in the park, and then at the train station, represented by the trajectory. The attacker could employ such sequence to retrieve the complete trajectory of $P$ in the released dataset: this attempt would succeed, provided that the attacker knows that $P$'s trajectory is actually present in the dataset, if the known trajectory is compatible with (i.e., is a sub-trajectory of) just one trajectory in the dataset. In this example of a linking attack in the trajectory domain, the sub-trajectory known by the attacker serves as QI, while the entire trajectory is the PI that is disclosed after the re-identification of the respondent. Clearly, as the example suggests, is rather difficult to distinguish QI and PI: in principle, any specific location can be the theater of a shadowing actions by a spy, and therefore any possible sequence of locations can be used as a QI, i.e., as a means for re-identification. As a consequence, distinguishing between QI and PI among the locations of a trajectory means putting artificial limits on the attacker's background knowledge; since most of the anonymity based techniques need assumption on the attacker's knowledge the approach is to make it as liberal as possible, in order to achieve maximal protection. In other words, the radical and typical assumption used is that any sub-trajectory that can be linked to a small number of individuals is a potentially dangerous QI and a potentially sensitive PI.

With this characteristic in mind, different papers proposed methods for making a trajectory dataset $k$-anonymous. Abul et al. [1] study the problem of privacy-preserving publishing of trajectory data and propose the notion of $(k, \delta)$-anonymity for movement data, where $\delta$ represents the possible location imprecision. In particular, this is a novel concept of $k$-anonymity based on co-localization that exploits the inherent uncertainty of the moving objects whereabouts. In this work authors also propose an approach, called *Never Walk Alone* (NWA), for obtaining a $(k, \delta)$-anonymous movement data. The method is based on trajectory clustering and spatial translation. In order to compute the trajectory similarity NWA applies Euclidean distance. This makes NWA only applicable to trajectories with equal length. Thus, the same authors in [2] propose W4M exploiting the EDR-based time tolerant distance. In particular, it groups $k$ similar trajectories by using a greedy clustering based on the EDR distance, and then exploits the minimum space translation via spatio-temporal editing to force all the trajectories of a cluster to be sufficiently similar with its center trajectory. Unfortunately, a successive work [107] proved that in general, these ap-

Table 3: Summary of methods based on ***Anonymity Models***.

| Name | Authors & Reference | Year | Privacy Model | Strategy | Data | |
|---|---|---|---|---|---|---|
| | | | | | Real | Synthetic |
| - | Terrovitis et al. [102] | 2008 | $k$-anonymity | suppression | | ✓ |
| NWA | Abul et al. [1] | 2008 | $(k, \delta)$-Anonymity | clustering, spatial translation | ✓ | ✓ |
| - | Nergiz et al. [77] | 2009 | $k$-Anonymity | clustering | ✓ | ✓ |
| - | Yarovoy et al. [114] | 2009 | $k$-Anonymity | generalization | ✓ | ✓ |
| KAM-REC | Monreale et al. [70] | 2010 | $k$-Anonymity | generalization | ✓ | |
| W4M | Abul et al. [2] | 2010 | $(k, \delta)$-Anonymity | clustering, spatial translation | ✓ | ✓ |
| Swap Location | Domingo-Ferrer et al. [26] | 2012 | $k$-Anonymity | clustering and permutation | ✓ | ✓ |
| ICBA | Gurung et al. [41] | 2014 | $k$-Anonymity | clustering | ✓ | ✓ |
| SeqAnon | Poulis et al. [85] | 2014 | $k^m$-anonymity | generalization | ✓ | ✓ |
| GLOVE | Gramaglia et al. [37] | 2015 | $k$-Anonymity | generalization | ✓ | |
| WCOP | Kopanaki et al. [53] | 2016 | Personalized $(k, \delta)$-Anonymity | clustering, spatial translation | ✓ | |
| SC-TDP | Li et al. [57] | 2016 | $k$-Anonymity | clustering and traj. segmentation | ✓ | ✓ |
| - | Mohammed et al. [69] | 2009 | $LKC$-privacy | global suppression | | |
| CAST | Monreale et al. [72] | 2011 | $c$-safety | semantic location generalization | ✓ | |
| - | Chen et al. [15] | 2013 | $LKC$-privacy | local suppression | ✓ | ✓ |
| SLAT | Liu et al. [61] | 2018 | $(\alpha, K, L)$-privacy | local suppression and attribute generalization | | ✓ |
| DPPP | Yao et al. [113] | 2019 | $(l, \alpha, \beta)$-privacy | location perturbation | ✓ | |
| - | Tu et al. [108] | 2019 | $k$-anonymity, $l$-diversity, $t$-closeness | generalization | ✓ | |

proaches offer trajectory $k$-anonymity only for $\delta = 0$, i.e., when each cluster contains at least $k$ identical trajectories. Clearly, in this case the uncertainty of trajectory is no longer exploited. In 2016, to improve the quality of the anonymized trajectories Li et al. [57] propose an alternative approach to obtain a $(k, \delta) - anonymity$ trajectory dataset. In particular, they introduce a segment clustering-based privacy-preserving algorithm which first divides the original data into blocks of similar trajectories and then, partitions trajectories belonging to each block into segments based on the minimum description length principle. The derived segments are anonymized by using a cluster-constraint strategy and guaranteeing that each clustering group satisfies $(k, \delta) - anonymity$.

Kopanaki et al. [53] extended the $(k, \delta)$-anonymization approaches to the aim to take into consideration different user privacy requirements. This approach first applies a trajectory segmentation to partition the trajectories into sub-trajectories. This transformation allows the clustering algorithm to discover similarities between the trajectories and to assign the respective partitions into clusters useful for applying the spatio-temporal translation and obtain the required anonymization. Since this setting requires personalized privacy requirements, the distortion applied in each cluster depends on the minimum value of $\delta$ required by the members of that cluster. The idea of personalized privacy in trajectory data has been introduced by Mahdavifar et al. [64] which considers each trajectory associated with the number of trajectories from which it should be indistinguishable (i.e., its privacy expectation). As a consequence the approach applies a clustering based approach that takes

into account similarities of trajectories also in terms of privacy expectation.

In [77] Nergiz et al. provide privacy protection in trajectory data by an approach based on two main steps: (1) first enforcing $k$-anonymity, meaning every released information refers to at least $k$ users/trajectories; and (2) then, reconstructing randomly a representation of the original dataset from the anonymization.

Another approach based on the concept of $k$-anonymity is proposed in [70], where Monreale et al. present a framework for $k$-anonymization of movement data combining the notions of spatial generalization and $k$-anonymity. Their approach is based on the idea to hide locations by means of generalization, specifically, replacing exact positions in the trajectories by approximate positions, i.e. points by centroids of areas obtained by spatial clustering of characteristic points of trajectories. The main steps involved in the proposed method are: *(i)* to construct a suitable tessellation of the geographical area into sub-areas in a data-driven fashion, i.e., that depends on the input trajectory dataset; *(ii)* to apply a spatial generalization of the original trajectories; *(iii)* to transform the dataset of generalized trajectories to ensure that it satisfies the notion of $k$-anonymity by exploiting a prefix-tree data structure representing the trajectories in a compact way.

Spatio-temporal generalization is taken into consideration to reach $k$-anonymity in [37], where the GLOVE approach is proposed. The basic idea of this approach is that most of the trajectories can be anonymized with limited loss of accuracy and only a smaller portion of trajectories requires drastic generalization. As a consequence they suggest a way to apply specialized generalization, i.e., each trajectory is affected by an independent and minimal reduction of granularity that hides it among other $k$-1 similar trajectories. The anonymization method computes for each pair of trajectories to be merged the cost in terms of loss of spatio-termporal granularity and iteratively, merges two trajectories with the smallest cost until each trajectory is $k$-anonymous.

Domingo-Ferrer and Trujillo-Rasua propose to reach the $k$-anonymity property in trajectory data by microaggregation and location permutations [26]. In particular, they group the trajectories into clusters of at least $k$ similar trajectories and then, permute the locations of the trajectories belonging to each cluster.

Poulis et al. [85] propose to adapt the notion of $k^m$-anonymity, introduced in [104] for transaction data publishing, to the mobility context. In particular, this privacy model assumes that an attacker has a background knowledge corresponding to a sub-trajectory of $m$ locations. The authors propose two anonymization algorithms applying generalization to increasingly larger parts of trajectories. Both algorithms do not consider the temporal information associated to each location. The SEQANON approach applies a distance-based generalization, creating generalized trajectories with locations that are close in proximity. It aims at preserving the distance between original locations. The SD-SEQANON approach considers the presence of a location taxonomy that enables a generalization that exploits the semantic similarity of trajectories. This algorithm tends to generalize trajectories whose locations are typically slightly more distant but much more semantically similar.

Most of these anonymization techniques do not take into consideration the actual road-network constraints. Gurung et al. [41] is one of the few works that propose a method that guarantees $k$-anonymity of trajectory data while generating trajectories following the road-network constraints. The anonymization approach is based on a trajectory clustering phase applied after eliminating records involving infrequent roads. The road-network constraints are especially enforced when computing the representative trajectory in a cluster.

All the anonymization techniques described so far are based on the assumption that it is hard to know for each trajectory data points representing QI for the individual. In the literature there exist methods, such as those introduced in [114, 102] that although deriving

spatio-temporal QI in the real-world is not an easy task, assume that QI may be provided directly by the users when they subscribe to the service, or be part of the users' personalized settings, or they may be found by means of statistical data analysis or data mining. Based on these assumptions these works designed anonymization algorithms that protect against adversaries exploiting the knowledge of QI. The approach presented in [114] is based on the idea that different individuals may have different QI, because in mobility it is not realistic to assume that a set o locations and time intervals can be QI for all the individuals in the data. As a consequence, anonymization groups associated with different individual trajectories may not be disjoint. Therefore, they introduce a variant of $k$-anonymity model that takes into consideration the possibility to re-identify individuals by combining different anonymization groups. In order to counter this type of privacy attacks and satisfy the $k$-anonymity property, Yarovoy et al. propose two approaches exploiting the space-generalization called *Extreme Union* and *Symmetric Anonymization* that differ in the strategy used to maintain under control the information loss that could derive from the overlapping of the anonymization groups which might force revisiting earlier spatial generalizations.

Terrovitis and Mamoulis [102] devise an anonymization approach that protects trajectory data against a specific number of adversaries that have the opportunity to reconstruct a subset of locations and link them to specific individual trajectories in the published data. In order to protect individuals against this set of adversaries they suggest an algorithm aiming at suppressing from the trajectories the dangerous data points. Since identifying the optimal set of points to be suppressed with the minimum possible information loss is an NP-hard problem, they propose a greedy heuristic solution that iteratively suppresses locations, until the privacy constraint is met. This approach simulates each possible attack that an adversary can conduct and then resolves the identified privacy breaches.

### 5.1.2 Approaches based on diversity of sensitive information

All the privacy techniques discussed in the previous section (Section 5.1.1) are able to counter the trajectory linking attack by releasing groups of trajectories with minimum size $k$ which are indistinguishable. Unfortunately, this property is not enough to counter attribute linking attacks that enables to derive sensitive information when individuals of the same anonymity group share similar values on some sensitive attributes. In the mobility literature there exist two different settings: protecting against the inference of sensitive information associated to each trajectory [69, 15, 52, 61, 113] and protecting against the inference of sensitive location data [72, 108].

The first category of approaches aims at defending the published records, composed by pairs of trajectories and sensitive attributes (e.g., disease) against two types of attacks: record linkage attack, which succeeds in case a trajectory in the database is so specific that not many individuals match the same data, and attribute linkage attack, which succeeds when a sensitive value occurs frequently with some trajectory and even if the record of an individual is not unique the adversary can infer the sensitive value.

Mohammed et al. [69] introduce a privacy model named $LKC$-privacy requiring that for each sub-trajectory with maximum length $L$ in a trajectory database exist at least $K$-1 trajectories and the confidence of inferring any sensitive attribute value is not greater than $C$. They also propose an anonymization technique that transforms the original database by global suppression of dangerous locations in order to enforce that requirement. In this context global suppression means that if a location is chosen to be suppressed, all instances of the location in the database are suppressed. This implies significant deterioration in the quality of the data. In order to improve the data quality in [15] Chen et al. some

years later propose a framework exploiting both local and global suppression to satisfy the $LKC$-privacy. Liu et al. [61] propose an extended version of the $LKC$-privacy model that considers also the protection of sensitive locations. The privacy model is called $(\alpha, K, L)$-privacy and requires that each sub-trajectory at most $L$ non-sensitive locations is shared by at least $K$ trajectories and the probability to infer any sensitive location or any sensitive value is less than $\alpha$. They also describe an algorithm for the enforcing of this privacy property called SLAT which combines trajectory splitting with location suppression and sensitive value generalization.

Another approach that is also able to prevent the similarity attack is that one presented in [113]. It protects the sensitive values by using perturbation, i.e., by adding or deleting some moving points and without changing any sensitive attribute. This approach guarantees a trajectory transformation that ensures the $(l, \alpha, \beta)$-privacy requiring $l$-diversity of the sensitive value, $\alpha$-sensitivity, i.e., the probability to infer sensitive value is less than $\alpha$, and $\beta$-similarity, i.e., the probability to infer that an individual possesses some sensitive value of a specific category is below $\beta$.

The second category of approaches [72, 108] considers the publication of trajectory data without any sensitive value assigned. Here, the only sensitive information to be protected is the sensitive locations. This case could be seen as a special case of the context presented by Liu et al. [61] where there is no sensitive attribute. Monreale et al. in [72] propose the privacy protection of sensitive locations of semantic trajectories, i.e., trajectories enriched with contextual information, by using a generalization techniques of the different visited places guided by a place taxonomy. The algorithm transforms the original trajectory dataset in a way that the $c$-safety property is satisfied; in particular this privacy model provides an upper bound $c$ to the probability of inferring that a given person, observed in a sequence of non-sensitive places, has also stopped in any sensitive location.

Lastly, Tu et al. [108] considers the setting where a trajectory is formed by a series of locations which may contain several PoIs, but some locations may have limited categories of PoIs or their distributions differ from the whole city. As a consequence, the authors propose an approach for the trajectory generalization that prevents both semantic attack and re-identification attack, by assuring $k$-anonymity, $l$-diversity and $t$-closeness at the same time. In particular, the proposed method during the merging operation of trajectories combines neighboring regions to make the resulting region satisfying $l$-diversity. This means that the number of distinct POI categories in that resulting region needs to exceed $l$. Similarly, in order to achieve $t$-closeness neighboring regions are merged until the divergence between its POI distribution and that of the entire city is no larger than the threshold $t$.

## 5.2 Privacy protection by Differential Privacy

In the mobility literature different approaches based on the differential privacy model [27] have been proposed. A differential private dataset can be obtained by using different mechanisms. The most used are: the Laplace mechanism, which adds Laplacian noise a vector of numerical values [27], and the exponential mechanism, which randomizes a probability distribution over a discrete, and finite set of values [66]. In both mechanisms the perturbation introduced in the data is calibrated by the parameter $\epsilon$ called privacy budget. The general idea of this privacy model is that regardless of the background knowledge, an adversary accessing to the private trajectory dataset comes to the same conclusion whether an individual's trajectory is included in the dataset or not. Thus, the differential privacy model can assure the individual that the released trajectory will not leak their privacy whether or not their trajectory is in the dataset.

Table 4: Summary of methods based on ***Differential Privacy***.

| Name | Authors & Reference | Year | Privacy Model | Mechanism | | Data | |
|---|---|---|---|---|---|---|---|
| | | | | Laplace | Exponential | Real | Synthetic |
| - | Chen et al. [14] | 2012 | $\epsilon$-differential privacy | ✓ | | ✓ | |
| $n$-grams | Chen et al. [13] | 2012 | $\epsilon$-differential privacy | ✓ | | ✓ | ✓ |
| SDD | Jiang et al. [49] | 2013 | $\epsilon$-differential privacy | | ✓ | ✓ | |
| DP-Where | Mir et al. [68] | 2013 | $\epsilon$-differential privacy | ✓ | ✓ | ✓ | |
| DPT | He et al. [42] | 2015 | $\epsilon$-differential privacy | ✓ | | ✓ | ✓ |
| - | Hua et al. [45] | 2015 | $\epsilon$-differential privacy | ✓ | ✓ | ✓ | |
| PrivTree | Zhang et al. [119] | 2016 | $\epsilon$-differential privacy | ✓ | | ✓ | |
| - | Li et al. [58] | 2017 | $\epsilon$-differential privacy | ✓ | ✓ | ✓ | |
| SafePath | Al-Hussaeni et al. [3] | 2018 | $\epsilon$-differential privacy | ✓ | | ✓ | |
| DP-Star | Gursoy et al. [40] | 2019 | $\epsilon$-differential privacy | ✓ | ✓ | ✓ | ✓ |
| NTPT | Zhao et al. [120] | 2020 | $\epsilon$-differential privacy | ✓ | | ✓ | |
| PDPDP | Tian et al. [105] | 2017 | PDP per trajectory | ✓ | | ✓ | |
| PLDP-TD | Deldar and Abadi [22] | 2018 | PDP per location | | | ✓ | ✓ |
| PDP-SAG | Deldar and Abadi [23] | 2019 | PDP per location | ✓ | | ✓ | ✓ |
| OPTDP | Cheng et al. [17] | 2022 | PDP per loc. and traj. | | | ✓ | |

### 5.2.1 Approaches exploiting Laplace Mechanism

Most of the approaches using the differential privacy model for guaranteeing privacy in trajectory data are based on the Laplace mechanism. Typically, they first transform the original trajectory exploiting a specific data representation that may be randomized by the differential privacy mechanism and then, the private data representation is used to generate the private set of trajectories. In general, these approaches aim at minimizing the noise added with the Laplace mechanism so that data utility is preserved as much as possible. Many of these works measure data utility by evaluating specific tasks after the protection mechanism is applied. Information loss therefore may vary depending on the type of data and approach.

This data representation enables the approximation of trajectory distribution in the dataset to be made private. Most of the solutions in the literature adopt a *hierarchical decomposition* approach, which recursively splits the trajectories into groups based on a similarity function and computes a noisy trajectory count for each group, until all noisy counts are less than a certain threshold. One of the first works based on this methodology is [14] that uses a prefix tree to represent the trajectory data with a hierarchical structure grouping trajectories having common prefix of location subsequences. The proposed method exploits a location taxonomy and the Laplace mechanism for efficiently constructing a noisy private prefix tree that considers multiple levels of spatial generalization. In particular, in each iteration, this approach first creates children nodes of the leaves of the previous iterations. These children correspond to the highest level of generalization of each potential location. Then, it adds Laplacian noise to the count of trajectories associated to each generalized node. Each node with a noisy count below a threshold is not expanded, while nodes with noisy counts above the threshold are expanded with nodes for all locations using the lower level of generalization. This process is repeated until a user-defined tree height is reached. Then the prefix tree is pruned for preserving only nodes representing the lowest level of generalization and the noisy counts of each node are made consistent across the levels, in order to guarantee that the count of each node is not less than the sum of counts of its children. As final step the resulting prefix tree is used to generate the differentially private trajectories. This approach is especially designed for sequence data, so sequence of

locations without any temporal information. However, the authors also describe how it is possible to extend the approach to trajectory data, where each trajectory is a sequence of pairs composed of (location, time). In this case, each node corresponds to a pair ($loc_i$, $t_i$) and then, for expanding such node in the prefix tree the combinations of all locations and the timestamps greater than $t_i$ should be taken into consideration. Unfortunately, the existence of the temporal dimension makes the algorithm not efficient due to the high-dimensionality and sparseness of data. This problem is addressed by Al-Hussaeni et al. [3] that propose an efficient and scalable differential private algorithm, called *SafePath*. As [14], they models trajectories in a prefix tree structure but use two different tree taxonomies: one for the location generalization and one for the time generalization. This is useful because helps in identifying empty nodes (i.e., nodes which are not representing any trajectory) so that they can filtered out as early as possible, for preventing false trajectories from being constructed. *SafePath* significantly reduces the runtime with its pruning strategy while data utility is greatly boosted.

A different tree structure is used by Chen et al. in [13] where first trajectory data are represented by $n-grams$ model corresponding to a Markov model of order $(n-1)$. In other words, each trajectory is described as transition probabilities based on a past history of $(n-1)$ locations. Based on the derived $n-gram$ representation of the trajectories the authors apply a tree construction procedure similar to that on described in [14], but without using any location taxonomy, so without any generalization level. In this case the Laplacian noise is added to the counts associated to the $n-grams$.

The limitation of these approaches is that they require a limit to the recursion depth in the splitting operation and that the noise to be added to counts has to be proportional to recursion depth. The choice of this parameter is problematic because it affects the noise and so, the data utility. To overcome this limitation Zhang et al. [119] propose an algorithm that also adopts the hierarchical approach but completely eliminates the dependency on a predefined recursion depth. The algorithm, called *PrivTree*, exploits a particular version of the Laplace mechanism which enables the use of only a constant amount of noise when decides whether a sub-domain should be split, without worrying about the recursion depth. The noise addition bound is possible because, as proved by Zhang et al., to publish a sequence one can add a noise amount that is not proportional to the sensitivity of that sequence.

The above approaches are based on the assumption that trajectories have a uniform speed over the time. Since, this assumption is not realistic, He et al. [42] propose an approach that exploits a set of hierarchically organized reference systems, derived by spatial discretization at different resolutions, to capture the fact that movements at slow speeds can be summarized using a fine granularity reference system, while movements with higher speeds are summarized using coarser granularity reference systems. The authors construct for each reference system a prefix tree that are perturbed by using the Laplace mechanism on the node counts. The standard process is followed with the only difference that the trajectories are synthetically generated by using a strategy that preserves the correct directionality in the output trajectories.

A prefix tree structure combined with the Laplace mechanism is also used in [120]. The main difference with respect to the previous works is that no spatial or temporal hierarchy is used and each node in the prefix tree represents the information about a segment of the trajectories and not a location. The approach takes into account also the presence of contextual information that could be used by an attacker and the differential privacy is used to randomize statistics like frequency and count of the segments.

### 5.2.2 Approaches exploiting the Exponential Mechanism

Although most of the approaches for privacy-preserving trajectory publication are based only on the Laplace mechanism, there exist different methods that also exploits the Exponential one [66].

As an example, Jiang et al. [49] present an algorithm that randomizes an input trajectory by sampling a suitable *distance* and *direction* at each position to publish the next possible position of the trajectory. These distance and direction values are sampled from exponential distributions guaranteeing strong differential privacy while maintaining a good quality of the trajectory data. Given a trajectory, this approach does not randomize the starting and ending point. Other approaches combine Laplace and Exponential mechanisms to generate synthetic private trajectories. Mir et al. [68], for instance, propose DP-WHERE which is based on a previous algorithm called WHERE [48], which is able to produce models describing how populations move within a metropolitan area. WHERE, starting from aggregations computed on human trajectory data, reproduces populations density over the time in a region. DP-WHERE can be seen as a variant of this algorithm that generates mobility traces from perturbed distributions of the extracted mobility features by mainly using the Laplace mechanism. The approach extracts and perturbs mobility features such as the probability distribution of *Home* cell over the grid cells of the territory, the probability distribution of *Work* cell over the grid cells, the *Commute Distance* distribution, i.e., distance between home and work, etc. While most of the features are perturbed by the Laplace noise, for making the *Commute Distance* distribution differentially private the authors propose a strategy that first, constructs the distribution by histogram bins guaranteeing differential privacy following the strategy already used in [20], that exploits the Exponential mechanism; and then, the histogram counts are perturbed by adding a Laplace noise.

Gursoy et al. [40] also combine the two mechanisms for achieving differential privacy with a completely different approach. They propose DP-STAR that is based on five main steps: one dedicated to pre-processing, four dedicated to preserve different types of spatial utility in trajectory data and one dedicated to the synthetic and private trajectory generation. In the pre-processing phase, DP-STAR represents each trajectory by its most *representative points*, that are able to summarize and characterize the overall movement. In the second step, the algorithm partitions the territory by using a density-aware adaptive grid structure, which is able to obtain small cells in high density regions and coarser cells in low density areas. The third phase is dedicated to the extraction of a trip distribution from the trajectory data; the forth step constructs a mobility model based on a first-order Markov model to have synthetic trajectories able to mimic the mobility patterns of actual trajectories; and the fifth phase tries to estimate the route length by using only the representative points contained in each trajectory. Finally, the last step generates the synthetic trajectory exploiting the features extracted in the previous steps. In order to guarantee the differential privacy DP-STAR uses the Laplace mechanism for the steps 2-4 while the Exponential mechanism is used for the route length estimation because it requires to compute the median length of the trajectories that start in a cell $c_i$ and end in a cell $c_j$. For computing this value Gursoy et al. propose a variant of the approach presented in [20].

A completely different schema for privacy-preserving trajectory publishing is presented in the works of Hua et al. and Li et al. [45, 58], where two differentially private algorithms are used: a differentially private $k$-means clustering is used for spatial generalization exploiting the exponential mechanism for sampling a clustering partition from an exponential distribution; and an algorithm for the trajectory publication based on random selection of generalized trajectories with locations drawn from generalized space and the addition of a

Laplace noise to the numbers of those trajectories.

### 5.2.3 Personalized Differential Privacy

Most of the privacy-preserving approaches developed for trajectory data assume that individuals have the same privacy expectation and thus, the algorithms are designed to guarantee the same level of privacy protection for all individuals. This assumption can lead to a situation where some individuals receive insufficient privacy and while other have an excess of privacy protection. To overcome this problem, in 2015 Jorgensen et al. [51] propose a new privacy model called Personalized Differential Privacy (PDP), which enables the specification of the privacy requirements at individual level. In the context of trajectory data only few works adopted this personalized approach even if it is promising especially from the utility viewpoint [105, 22, 23, 17] .

Tian et al. [105] assume that in the dataset there is a trajectory for each user who has the own individual privacy preference for the whole trajectory. In order to guarantee the privacy protection required by each individual they propose an approach based on two main steps: trajectories generalization and PDP trajectories generation. The trajectory generalization is obtained by applying a Hilbert Curve based location clustering approach and using the clustering result for the spatial generalization of each timestamp. Since we are in a setting where each user has a different privacy requirement then, the centroid cannot be used to represent the locations belonging to the same location cluster. Thus, the authors propose an approach to compute the representative element of a cluster on the basis of the contribution of the different users which varies with their privacy expectation. The proposed strategy assures that the location belonging to the conservative user will contribute less in the final representative element than the location belonging to the more liberal user. The location generalization leads to have a decrease of the distinguished locations at each timestamp and thus, it will help the compact representation of the trajectory data by using a prefix tree structure that assumes common prefixes of the trajectories under analysis. The prefix tree is made differentially private by adding Laplacian noise to the different nodes and it is used for the trajectory generation.

This approach assumes the same privacy requirement for each location. Deldar and Abadi [22] propose an approach to construct a personalized noisy trajectory tree based on the underlying trajectory database and different privacy protection requirements of the involved locations. Each node in the tree represents a sub-trajectory and a personalized noise is added to its count. In this work the privacy requirement does not depend on the user expectation but on the location itself. After the construction of the noisy trajectory tree, Deldar and Abadi enforce some consistency constraints to guarantee that the noisy count of each non-leaf node should be equal to the sum of its children's noisy counts. The same authors address also the problem to satisfy personalized differential privacy of trajectory data in case it is enriched with sensitive non-spatiotemporal attributes [23]. To solve the problem they consider for each node of the tree representing a subtrajectory also a taxonomy tree for the generalization of the sensitive attribute to be generalized according to the privacy requirement expressed by the trajectory's user.

The above works set different privacy level for different users or different locations, but they cannot achieve privacy protection on both user and location level. Moreover, they require to set the privacy requirement as input. These two issues have been recently addressed by Cheng et al. [17] that propose an approach that is able to assign to each user location neither the privacy requirement derived in a data-drive fashion or defined by the user. For automatically deriving the privacy needs, the approach takes into consideration

Table 5: Summary of methods based on ***Generative Models***.

| Name | Authors & Reference | Year | Generative Model | Data | |
|---|---|---|---|---|---|
| | | | | Real | Synthetic |
| - | Kulkarni et al. [55] | 2018 | SGAN,RGAN, RNN-LSTM | ✓ | |
| TrajGAN | Liu et al. [60] | 2018 | GAN | | |
| - | Yin and Yang in [115] | 2018 | GAN | ✓ | |
| LSTM-TrajGAN | Rao et al. [91] | 2021 | LSTM-GAN | ✓ | |
| LSTM-PAE | Zhan et al [118] | 2022 | LSTM Auto-Encoder | ✓ | |

the type of user location and assigns a different protection need for example to a location that is a stay-point of a frequent sub-trajectory with respect to a location that is stay-point of an infrequent sub-trajectory. In particular, Cheng et al. propose an approach that first builds a probabilistic mobility model for trajectories, i.e., a time-dependent first-order Markov chain on the set of locations. Second, it applies a clustering of the locations on different trajectories based on the mobility model and a *semantic similarity* function, and get the best semantic location matching results between different trajectories and the semantic similarity under this matching. Given the result of this matching, it extracts the most representative template trajectories according to the semantic similarity. The template trajectories receive a privacy level that depend on their type of locations. Finally, according to the matching results of template trajectory and other trajectories, the privacy levels of all trajectory locations are obtained. Once defined the privacy level of all trajectory based on their similarity with specific templates, the corresponding privacy budget is allocated and the final publishable trajectory data is obtained.

## 5.3 Privacy by generating synthetic trajectories

One of the main issues that we need to address during the design and the application of any privacy-preserving technique on trajectory data as well as on any other type of data is the trade-off between protection and data utility. This aspect is really hard to control: a very good approach may obscure the trajectory data perfectly protecting the spatio-temporal privacy of users, but cannot ensure the data quality. This is a great issue because data becomes useless from the analytical viewpoint.

As alternative to the classical approaches for guaranteeing privacy, some researchers propose to exploit the advancements in machine learning to develop a new family of privacy protection techniques for trajectory data using generative models based on deep learning [115, 55, 60, 91, 118]. In particular, most of them suggest to use generative model based on Generative Adversarial Networks (GANs [36]), i.e., neural network models able to generate high-quality synthetic data which follow the same distribution of training data. Typically, a GAN is composed of two neural networks: a generator and a discriminator. Their combination is able to learn the original data distribution by playing a minimax game. The trajectory discriminator has the goal to understand whether the trajectory samples is a real trajectory or is synthetically generated. Thus, the goal of the trajectory generator is to generate high-quality synthetic trajectories that can fool the trajectory discriminator.

Kulkarni et al. [55] presents the first experimental analysis on the use of different deep learning models for synthetically generating trajectory data. The experiments evaluate different factors such as the time required by each model, and more important the privacy-quality trade-off of the generated trajectories. To asses the privacy this work considers a location-sequence attack [95], and a membership interference attack [90]. This work compares different network architectures based on recurrent neural networks, such as Char-

RNN [39], RNN-LSTM [44], and recurrent highway networks [123], and two based on GANs: SGAN [116] and RGAN [31].

An experimental analysis of the privacy guarantee of generative models is also presented by Yin and Yang in [115]. They propose to use a GAN to generate a synthetic location density matrix with a better privacy-reality trade-off than that of the existing noise based approaches for differential privacy [5]. Note that, a location density matrix is an aggregate version of mobility data that is used instead of individual trajectories. The idea of using a GAN for protecting privacy is based on the fact that the generator introduces some noise during the generation, thus it is possible to use that noise instead of a differential privacy approach. Clearly, the privacy protection directly depends on the training goodness: a better data generation implies lower privacy. As a consequence, one should determine the appropriate number for the training iterations to balance this privacy-quality trade-off.

Liu et al. [60] propose TrajGAN, a theoretical framework based on a GAN model for synthetically generating human trajectories and discusses the possible challenges that can derive from using this kind of approach. For example, possible drawbacks could be the loss of occasional travels of people or privacy violation in case of overfitting of the model which thus is not able to generalize high-level patterns in trajectories. Rao et al. [91], inspired by the vision of the TrajGANs [60], propose an implementation of that theoretical framework called LSTM-TrajGAN. This approach consists of three main components: *(i)* a Trajectory Encoding Model, which encodes location coordinates, temporal attributes, and other attributes such as point of interest (POI) category; *(ii)* a Trajectory Generator, which takes random noise and original trajectories as inputs to generate synthetic trajectories; and *(iii)* a Trajectory Discriminator, which takes trajectories as inputs and determines them as *real* or *synthetic*. Both the Trajectory Generator and Discriminator are based on the Long Short-Term Memory model suitable for data with sequential nature.

Recently, Zhan et al. [118] proposed an alternative to the use of GANs for guaranteeing privacy in trajectory data. In particular, the proposed approach exploits the adversarial learning to better balance the potential trade-off between privacy and utility. It is based on an LSTM auto-encoder with three main components: (i) a Mobility Prediction Unit that takes as input the trajectory data and optimizes the prediction task representing the means for measuring utility; (ii) a User Re-identification Risk Unit, which is a neural network that solves the task to re-identifying the user of a trajectory; (iii) Data Reconstruction Risk Unit which evaluate the differences between the reconstructed trajectory and the original input trajectory.

# 6 Conclusion

Privacy issues in mobility and location data are recognized as an important and challenging problem from both the scientific and legal standpoint. Given the widespread development and adoption of location-based applications and technologies, the research concerning techniques for an empirical assessment and mitigation of privacy risk is of the utmost importance. This survey discussed the advancement of the scientific literature in privacy-preserving mobility data publication, focusing on adversary attacks, privacy models, privacy risk assessment techniques and privacy protection and mitigation algorithms. There are several interesting open directions for future research in this area. All the privacy-preserving techniques reviewed in this survey in the design of the mitigation strategy assume a worst-case scenario for the adversary attack models. This often leads to the very hard challenge to maintain under control the utility of the mobility data under analysis.

Unfortunately, most of the time the theoretical attack model is not realistic because assumes for example a knowledge in posses of the attacker that is very hard to collect. This problem has been highlighted in [84] that proved that by considering an adversary who moves on the territory to collect the information for attacking the target respecting spatio-temporal environmental constraints, the risk caused by such adversary is lower than the theoretical worst-case adversary model. This considerations suggest that, in order to improve the trade-off between privacy and data utility, it would be interesting to design mitigation strategies that take into consideration more realistic adversarial attacks and that use the information about the empirical risk evaluation that can be provided by privacy risk assessment frameworks. An approach that focus on mitigating the actual privacy risk could be beneficial for the data quality while assuring privacy protection.

The combined use of privacy risk assessment and mitigation strategies could also help the research on personalized privacy. Current research works assume that the privacy preferences are defined by the individuals through a specific function, on the basis of the type of locations or by patterns matching some templates recognized as risky. The use of privacy risk assessment tools could help in identifying with a data-driven approach the actual risky locations for an individual, assigning to each location or sub-trajectory the corresponding risk. In other words, the personalization of the privacy preferences could be determined looking at the actual risk. Analyzing the literature it is evident that the study of personalized privacy in the context of mobility data is an open and very promising field. Most of the works are based on the personalized differential privacy and only very few works consider the personalization principle for other privacy models. Although determining the privacy budget is recognized as a hard task even for privacy experts, none of these works addresses the problem of how individuals can express the privacy budget corresponding to the their privacy expectation with some level of awareness. This particular aspects represents a gap of fundamental importance for giving to individuals the ability to share their data with awareness and control. Finally, even though research proposing the use of synthetic trajectories for privacy protection is promising and interesting, the current approaches are mostly based on empirical evidence on the privacy guaranteed. Future research in this context should focus also on the development of strategies that by design incorporate a privacy mechanism more controllable from the onset and formally provable.

**Acknowledgements** This work is partially supported by the European Community H2020 programme under the funding schemes: H2020-INFRAIA-2019-1: Research Infrastructure G.A. 871042 SoBigData++ (sobigdata.eu), G.A. 952215 TAILOR and G.A. 952026 Humane AI NET (humane-ai.eu)

# References

[1] O. Abul, F. Bonchi, and M. Nanni. Never walk alone: Uncertainty for anonymity in moving objects databases. In *ICDE*, pages 376–385. IEEE Computer Society, 2008.

[2] O. Abul, F. Bonchi, and M. Nanni. Anonymization of moving objects databases by clustering and perturbation. *Inf. Syst.*, 35(8):884–910, 2010.

[3] K. Al-Hussaeni, B. C. M. Fung, F. Iqbal, G. G. Dagher, and E. G. Park. Safepath: Differentially-private publishing of passenger trajectories in transportation systems. *Comput. Networks*, 143:126–139, 2018.

[4] L. Alessandretti. What human mobility data tell us about covid-19 spread. *Nature Reviews Physics*, 4(1):12–13, 2022.

[5] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Geo-indistinguishability: differential privacy for location-based systems. In *CCS*, pages 901–914. ACM, 2013.

[6] M. Atzori, F. Bonchi, F. Giannotti, and D. Pedreschi. Anonymity preserving pattern discovery. *VLDB J.*, 17(4):703–727, 2008.

[7] M. Backes and S. Meiser. Differentially private smart metering with battery recharging. In *DPM/SETOP*, volume 8247 of *Lecture Notes in Computer Science*, pages 194–212. Springer, 2013.

[8] F. Basïk, H. Ferhatosmanoğlu, and B. Gedik. Slim: Scalable linkage of mobility data. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, SIGMOD '20, page 1181–1196, New York, NY, USA, 2020. Association for Computing Machinery.

[9] C. Bettini. Privacy protection in location-based services: A survey. In *Handbook of Mobile Data Privacy*, pages 73–96. Springer, 2018.

[10] F. Bonchi, L. V. S. Lakshmanan, and W. H. Wang. Trajectory anonymity in publishing personal mobility data. *SIGKDD Explor.*, 13(1):30–42, 2011.

[11] A. Cavoukian. Privacy design principles for an integrated justice system - working paper. 2000. urlhttps://www.ipc.on.ca/english/Resources/Discussion-Papers/Discussion-Papers-Summary/?id=318.

[12] A. Cavoukian. Privacy by design the 7 foundational principles. August 2009.

[13] R. Chen, G. Ács, and C. Castelluccia. Differentially private sequential data publication via variable-length n-grams. In *CCS*, pages 638–649. ACM, 2012.

[14] R. Chen, B. C. M. Fung, B. C. Desai, and N. M. Sossou. Differentially private transit data publication: a case study on the montreal transportation system. In *KDD*, pages 213–221. ACM, 2012.

[15] R. Chen, B. C. M. Fung, N. Mohammed, B. C. Desai, and K. Wang. Privacy-preserving trajectory data publishing by local suppression. *Inf. Sci.*, 231:83–97, 2013.

[16] X. Chen, A. Mizera, and J. Pang. Activity tracking: A new attack on location privacy. In *2015 IEEE Conference on Communications and Network Security (CNS)*, pages 22–30, 2015.

[17] W. Cheng, R. Wen, H. Huang, W. Miao, and C. Wang. OPTDP: towards optimal personalized trajectory differential privacy for trajectory data publishing. *Neurocomputing*, 472:201–211, 2022.

[18] P. Cintia, D. Fadda, F. Giannotti, L. Pappalardo, G. Rossetti, D. Pedreschi, S. Rinzivillo, P. Bonato, F. Fabbri, F. Penone, M. Savarese, D. Checchi, F. Chiaromonte, P. Vineis, G. Guzzetta, F. Riccardo, V. Marziano, P. Poletti, F. Trentini, A. Bella, X. Andrianou, M. D. Manso, M. Fabiani, S. Bellino, S. Boros, A. M. Urdiales, M. F. Vescio, S. Brusaferro, G. Rezza, P. Pezzotti, M. Ajelli, and S. Merler. The relationship between human mobility and viral transmissibility during the COVID-19 epidemics in italy. *CoRR*, abs/2006.03141, 2020.

[19] CISCO. Cisco consumer privacy survey 2019. url=http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

[20] G. Cormode, C. M. Procopiuc, D. Srivastava, E. Shen, and T. Yu. Differentially private spatial decompositions. In *ICDE*, pages 20–31. IEEE Computer Society, 2012.

[21] Y.-A. De Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3(1):1–5, 2013.

[22] F. Deldar and M. Abadi. PLDP-TD: personalized-location differentially private data analysis on trajectory databases. *Pervasive Mob. Comput.*, 49:1–22, 2018.

[23] F. Deldar and M. Abadi. PDP-SAG: personalized privacy protection in moving objects databases by combining differential privacy and sensitive attribute generalization. *IEEE Access*, 7:85887–85902, 2019.

[24] D. do Couto Teixeira, J. M. Almeida, and A. C. Viana. On estimating the predictability of

human mobility: the role of routine. *EPJ Data Sci.*, 10(1):49, 2021.

[25] J. Domingo-Ferrer and V. Torra. Ordinal, continuous and heterogeneous *k*-anonymity through microaggregation. *Data Min. Knowl. Discov.*, 11(2):195–212, 2005.

[26] J. Domingo-Ferrer and R. Trujillo-Rasua. Microaggregation- and permutation-based anonymization of movement data. *Inf. Sci.*, 208:55–80, 2012.

[27] C. Dwork. Differential privacy. In *ICALP (2)*, volume 4052 of *Lecture Notes in Computer Science*, pages 1–12. Springer, 2006.

[28] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC '06*, pages 265–284, 2006.

[29] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Differential privacy – a primer for the perplexed. 2011.

[30] Enisa. Ccta risk analysis and management method. url=https://www.enisa.europa.eu/topics/threat-risk-management/risk-management/current-risk/risk-management-inventory/rm-ra-methods/m_cramm.html.

[31] C. Esteban, S. L. Hyland, and G. Rätsch. Real-valued (medical) time series generation with recurrent conditional gans. *CoRR*, abs/1706.02633, 2017.

[32] M. Fiore, P. Katsikouli, E. Zavou, M. Cunche, F. Fessant, D. L. Hello, U. M. Aïvodji, B. Olivier, T. Quertier, and R. Stanica. Privacy in trajectory micro-data publishing: a survey. *Trans. Data Priv.*, 13(2):91–149, 2020.

[33] J. Freudiger, M. H. Manshaei, J.-P. Hubaux, and D. C. Parkes. On non-cooperative location privacy: A game-theoretic analysis. In *Proceedings of the 16th ACM Conference on Computer and Communications Security*, CCS '09, page 324–337, New York, NY, USA, 2009. Association for Computing Machinery.

[34] J. Freudiger, R. Shokri, and J. Hubaux. Evaluating the privacy risk of location-based services. In *Financial Cryptography*, volume 7035 of *Lecture Notes in Computer Science*, pages 31–46. Springer, 2011.

[35] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu. Privacy-preserving data publishing: A survey of recent developments. *ACM Comput. Surv.*, 42(4), jun 2010.

[36] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio. Generative adversarial networks. *Commun. ACM*, 63(11):139–144, 2020.

[37] M. Gramaglia and M. Fiore. Hiding mobile traffic fingerprints with GLOVE. In *CoNEXT*, pages 26:1–26:13. ACM, 2015.

[38] M. Gramaglia, M. Fiore, A. Tarable, and A. Banchs. Preserving mobile subscriber privacy in open datasets of spatiotemporal trajectories. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pages 1–9, 2017.

[39] S. Grossberg. Recurrent neural networks. *Scholarpedia*, 8(2):1888, 2013.

[40] M. E. Gursoy, L. Liu, S. Truex, and L. Yu. Differentially private and utility preserving publication of trajectory data. *IEEE Trans. Mob. Comput.*, 18(10):2315–2329, 2019.

[41] S. Gurung, D. Lin, W. Jiang, A. R. Hurson, and R. Zhang. Traffic information publication with privacy preservation. *ACM Trans. Intell. Syst. Technol.*, 5(3):44:1–44:26, 2014.

[42] X. He, G. Cormode, A. Machanavajjhala, C. M. Procopiuc, and D. Srivastava. DPT: differentially private trajectory synthesis using hierarchical reference systems. *Proc. VLDB Endow.*, 8(11):1154–1165, 2015.

[43] M. Henriquez. The top data breaches of 2021. *Security Magazine - BNP Media Security Group*, December 2021.

[44] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, 1997.

[45] J. Hua, Y. Gao, and S. Zhong. Differentially private publication of general time-serial trajectory data. In *INFOCOM*, pages 549–557. IEEE, 2015.

[46] S. Ibrahim and A. M. Omer. Survey on k-anonymity: Methods based on generalization technique. In *ICISA*, pages 1–2. IEEE, 2013.

[47] N. I. o. S. Information Technology Laboratory and Technology. Risk management guide for information technology systems, special publication 800-30. url=http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf.

[48] S. Isaacman, R. A. Becker, R. Cáceres, M. Martonosi, J. Rowland, A. Varshavsky, and W. Willinger. Human mobility modeling at metropolitan scales. In *MobiSys*, pages 239–252. ACM, 2012.

[49] K. Jiang, D. Shao, S. Bressan, T. Kister, and K. Tan. Publishing trajectories with differential privacy guarantees. In *SSDBM*, pages 12:1–12:12. ACM, 2013.

[50] F. Jin, W. Hua, M. Francia, P. Chao, M. Orlowska, and X. Zhou. A Survey and Experimental Study on Privacy-Preserving Trajectory Data Publishing. 1 2021.

[51] Z. Jorgensen, T. Yu, and G. Cormode. Conservative or liberal? personalized differential privacy. In *ICDE*, pages 1023–1034. IEEE Computer Society, 2015.

[52] E. G. Komishani, M. Abadi, and F. Deldar. PPTD: preserving personalized privacy in trajectory data publishing by sensitive attribute generalization and trajectory local suppression. *Knowl. Based Syst.*, 94:43–59, 2016.

[53] D. Kopanaki, V. Theodossopoulos, N. Pelekis, I. Kopanakis, and Y. Theodoridis. Who cares about others' privacy: Personalized anonymization of moving object trajectories. In *EDBT*, pages 425–436. OpenProceedings.org, 2016.

[54] J. Krumm. Inference attacks on location tracks. In *Pervasive*, volume 4480 of *Lecture Notes in Computer Science*, pages 127–143. Springer, 2007.

[55] V. Kulkarni, N. Tagasovska, T. Vatter, and B. Garbinato. Generative models for simulating mobility trajectories. *CoRR*, abs/1811.12801, 2018.

[56] K. LeFevre, D. DeWitt, and R. Ramakrishnan. Mondrian Multidimensional K-Anonymity. In *22nd International Conference on Data Engineering (ICDE'06)*, pages 25–25, Atlanta, GA, USA, 2006. IEEE.

[57] F. Li, F. Gao, L. Yao, and Y. Pan. Privacy preserving in the publication of large-scale trajectory databases. In *BigCom*, volume 9784 of *Lecture Notes in Computer Science*, pages 367–376. Springer, 2016.

[58] M. Li, L. Zhu, Z. Zhang, and R. Xu. Achieving differential privacy of trajectory data publishing in participatory sensing. *Inf. Sci.*, 400:1–13, 2017.

[59] N. Li, T. Li, and S. Venkatasubramanian. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In *2007 IEEE 23rd International Conference on Data Engineering*, pages 106–115, Istanbul, Apr. 2007. IEEE.

[60] X. Liu, H. Chen, and C. Andris. trajgans: Using generative adversarial networks for geo-privacy protection of trajectory data (vision paper). In *Location Privacy and Security Workshop*, pages 1–7, 2018.

[61] X. Liu, L. Wang, and Y. Zhu. SLAT: sub-trajectory linkage attack tolerance framework for privacy-preserving trajectory publishing. In *NaNA*, pages 298–303. IEEE, 2018.

[62] M. Luca, G. Barlacchi, B. Lepri, and L. Pappalardo. A survey on deep learning for human mobility. *ACM Computing Surveys (CSUR)*, 55(1):1–44, 2021.

[63] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam. L-diversity: privacy beyond k-anonymity. In *22nd International Conference on Data Engineering (ICDE'06)*, pages 24–24, Atlanta, GA, USA, 2006. IEEE.

[64] S. Mahdavifar, M. Abadi, M. Kahani, and H. Mahdikhani. A clustering-based approach for

personalized privacy preserving publication of moving object trajectory data. In *NSS*, volume 7645 of *Lecture Notes in Computer Science*, pages 149–165. Springer, 2012.

[65] E.-L. Makri, Z. Georgiopoulou, and C. Lambrinoudakis. A proposed privacy impact assessment method using metrics based on organizational characteristics. In S. Katsikas, F. Cuppens, N. Cuppens, C. Lambrinoudakis, C. Kalloniatis, J. Mylopoulos, A. Antón, S. Gritzalis, F. Pallas, J. Pohle, A. Sasse, W. Meng, S. Furnell, and J. Garcia-Alfaro, editors, *Computer Security*, pages 122–139, Cham, 2020. Springer International Publishing.

[66] F. McSherry and K. Talwar. Mechanism design via differential privacy. In *FOCS*, pages 94–103. IEEE Computer Society, 2007.

[67] A. Meyerson and R. Williams. On the complexity of optimal K-anonymity. In *Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems - PODS '04*, page 223, Paris, France, 2004. ACM Press.

[68] D. J. Mir, S. Isaacman, R. Cáceres, M. Martonosi, and R. N. Wright. DP-WHERE: differentially private modeling of human mobility. In *IEEE BigData*, pages 580–588. IEEE Computer Society, 2013.

[69] N. Mohammed, B. C. M. Fung, and M. Debbabi. Walking in the crowd: anonymizing trajectory data for pattern analysis. In *CIKM*, pages 1441–1444. ACM, 2009.

[70] A. Monreale, G. L. Andrienko, N. V. Andrienko, F. Giannotti, D. Pedreschi, S. Rinzivillo, and S. Wrobel. Movement data anonymity through generalization. *Trans. Data Priv.*, 3(2):91–121, 2010.

[71] A. Monreale, D. Pedreschi, R. G. Pensa, and F. Pinelli. Anonymity preserving sequential pattern mining. *Artif. Intell. Law*, 22(2):141–173, 2014.

[72] A. Monreale, R. Trasarti, D. Pedreschi, C. Renso, and V. Bogorny. C-safety: a framework for the anonymization of semantic trajectories. *Trans. Data Priv.*, 4(2):73–101, 2011.

[73] E. Murati and M. Henkoja. Location data privacy on maas under gdpr. *Eur. J. Privacy L. & Tech.*, page 115, 2019.

[74] M. Nanni, G. L. Andrienko, A. Barabási, C. Boldrini, F. Bonchi, C. Cattuto, F. Chiaromonte, G. Comandé, M. Conti, M. Coté, F. Dignum, V. Dignum, J. Domingo-Ferrer, P. Ferragina, F. Giannotti, R. Guidotti, D. Helbing, K. Kaski, J. Kertész, S. Lehmann, B. Lepri, P. Lukowicz, S. Matwin, D. Megías, A. Monreale, K. Morik, N. Oliver, A. Passarella, A. Passerini, D. Pedreschi, A. Pentland, F. Pianesi, F. Pratesi, S. Rinzivillo, S. Ruggieri, A. Siebes, V. Torra, R. Trasarti, J. van den Hoven, and A. Vespignani. Give more data, awareness and control to individual citizens, and they will help COVID-19 containment. *Trans. Data Priv.*, 13(1):61–66, 2020.

[75] F. Naretto, R. Pellungrini, A. Monreale, F. M. Nardini, and M. Musolesi. Predicting and explaining privacy risk exposure in mobility data. In A. Appice, G. Tsoumakas, Y. Manolopoulos, and S. Matwin, editors, *Discovery Science - 23rd International Conference, DS 2020*, volume 12323 of *Lecture Notes in Computer Science*, pages 403–418. Springer, 2020.

[76] F. Naretto, R. Pellungrini, F. M. Nardini, and F. Giannotti. Prediction and explanation of privacy risk on mobility data with neural networks. In *PKDD/ECML Workshops*, volume 1323 of *Communications in Computer and Information Science*, pages 501–516. Springer, 2020.

[77] M. E. Nergiz, M. Atzori, Y. Saygin, and B. Güç. Towards trajectory anonymization: a generalization-based approach. *Trans. Data Priv.*, 2(1):47–75, 2009.

[78] M. Nyhan, I. Kloog, R. Britter, C. Ratti, and P. Koutrakis. Quantifying population exposure to air pollution using individual mobility patterns inferred from mobile phone data. *Journal of exposure science & environmental epidemiology*, 29(2):238–247, 2019.

[79] OWASP. Risk rating methodology. url=http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

[80] L. Pappalardo, F. Simini, S. Rinzivillo, D. Pedreschi, F. Giannotti, and A.-L. Barabási. Returners

and explorers dichotomy in human mobility. *Nature communications*, 6(1):1–8, 2015.

[81] L. Pappalardo, M. Vanhoof, L. Gabrielli, Z. Smoreda, D. Pedreschi, and F. Giannotti. An analytical framework to nowcast well-being using mobile phone data. *Int. J. Data Sci. Anal.*, 2(1-2):75–92, 2016.

[82] E. Parliament. General data protection regulation. url=http://data.europa.eu/eli/reg/2016/679/oj.

[83] R. Pellungrini, L. Pappalardo, F. Pratesi, and A. Monreale. A data mining approach to assess privacy risk in human mobility data. *ACM Trans. Intell. Syst. Technol.*, 9(3), dec 2017.

[84] R. Pellungrini, L. Pappalardo, F. Simini, and A. Monreale. Modeling adversarial behavior against mobility data privacy. *IEEE Trans. Intell. Transp. Syst.*, 23(2):1145–1158, 2022.

[85] G. Poulis, S. Skiadopoulos, G. Loukides, and A. Gkoulalas-Divanis. Apriori-based algorithms for $k^m$-anonymizing trajectory data. *Trans. Data Priv.*, 7(2):165–194, 2014.

[86] F. Prasser, J. Eicher, H. Spengler, R. Bild, and K. A. Kuhn. Flexible data anonymization using arx—current status and challenges ahead. *Software: Practice and Experience*, 50:1277 – 1304, 2020.

[87] F. Prasser, F. Kohlmayer, R. Lautenschläger, and K. A. Kuhn. Arx–a comprehensive tool for anonymizing biomedical data. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2014:984–993, 11 2014.

[88] F. Pratesi, A. Monreale, R. Trasarti, F. Giannotti, D. Pedreschi, and T. Yanagihara. Prudence: A system for assessing privacy risk vs utility in data sharing ecosystems. *Transactions on Data Privacy*, 11:139–167, 08 2018.

[89] W. Presthus and H. Sørum. Are consumers concerned about privacy? an online survey emphasizing the general data protection regulation. *Procedia Computer Science*, 138:603–611, 2018. CENTERIS 2018 - International Conference on ENTERprise Information Systems / ProjMAN 2018 - International Conference on Project MANagement / HCist 2018 - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN/HCist 2018.

[90] A. Pyrgelis, C. Troncoso, and E. D. Cristofaro. Knock knock, who's there? membership inference on aggregate location data. In *NDSS*. The Internet Society, 2018.

[91] J. Rao, S. Gao, Y. Kang, and Q. Huang. Lstm-trajgan: A deep learning approach to trajectory privacy protection. In *GIScience (I)*, volume 177 of *LIPIcs*, pages 12:1–12:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.

[92] L. Rossi and M. Musolesi. It's the way you check-in: identifying users in location-based social networks. In *COSN '14*, 2014.

[93] P. Samarati and L. Sweeney. Generalizing data to provide anonymity when disclosing information (abstract). In *PODS*, page 188. ACM Press, 1998.

[94] T. Shevlane, B. Garfinkel, and A. Dafoe. Contact tracing apps can help stop coronavirus. but they can hurt privacy. *The Washington Post*, April 2020.

[95] R. Shokri, G. Theodorakopoulos, J. L. Boudec, and J. Hubaux. Quantifying location privacy. In *IEEE Symposium on Security and Privacy*, pages 247–262. IEEE Computer Society, 2011.

[96] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.

[97] X. Song, Q. Zhang, Y. Sekimoto, R. Shibasaki, N. J. Yuan, and X. Xie. Prediction and simulation of human mobility following natural disasters. *ACM Trans. Intell. Syst. Technol.*, 8(2):29:1–29:23, 2017.

[98] M. Srivatsa and M. Hicks. Deanonymizing mobility traces: Using social network as a side-channel. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS '12, page 628–637, New York, NY, USA, 2012. Association for Computing Machinery.

[99] K. Stokes and V. Torra. Reidentification and k-anonymity: a model for disclosure risk in

graphs. *Soft Comput.*, 16(10):1657–1670, 2012.

[100] K. Sui, Y. Zhao, D. Liu, M. Ma, L. Xu, Z. Li, and D. Pei. Your trajectory privacy can be breached even if you walk in groups. *2016 IEEE/ACM 24th International Symposium on Quality of Service (IWQoS)*, pages 1–6, 2016.

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

[102] M. Terrovitis and N. Mamoulis. Privacy preservation in the publication of trajectories. In *MDM*, pages 65–72. IEEE, 2008.

[103] M. Terrovitis, N. Mamoulis, and P. Kalnis. Privacy-preserving anonymization of set-valued data. *Proc. VLDB Endow.*, 1(1):115–125, 2008.

[104] M. Terrovitis, N. Mamoulis, and P. Kalnis. Local and global recoding methods for anonymizing set-valued data. *VLDB J.*, 20(1):83–106, 2011.

[105] F. Tian, S. Zhang, L. Lu, H. Liu, and X. Gui. A novel personalized differential privacy mechanism for trajectory data publication. In *NaNA*, pages 61–68. IEEE Computer Society, 2017.

[106] V. Torra. *Data Privacy: Foundations, New Developments and the Big Data Challenge*. Springer Publishing Company, Incorporated, 1st edition, 2017.

[107] R. Trujillo-Rasua and J. Domingo-Ferrer. On the privacy offered by (k, $\delta$)-anonymity. *Inf. Syst.*, 38(4):491–494, 2013.

[108] Z. Tu, K. Zhao, F. Xu, Y. Li, L. Su, and D. Jin. Protecting trajectory from semantic attack considering ${k}$ -anonymity, ${l}$ -diversity, and ${t}$ -closeness. *IEEE Trans. Netw. Serv. Manag.*, 16(1):264–278, 2019.

[109] J. Valentino-DeVries, N. S. M. H. Keller, and A. Krolik. Your apps know where you were last night, and they're not keeping it secret. *The New York Times*, December 2018.

[110] I. Wagner and D. Eckhoff. Privacy Assessment in Vehicular Networks Using Simulation. In *Winter Simulation Conference (WSC '14)*, pages 3155–3166, Savannah, GA, December 2014. IEEE.

[111] I. Wagner and D. Eckhoff. Technical privacy metrics: A systematic survey. *ACM Comput. Surv.*, 51(3):57:1–57:38, 2018.

[112] Y. Xu, B. C. M. Fung, K. Wang, A. W. Fu, and J. Pei. Publishing sensitive transactions for itemset utility. In *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008), December 15-19, 2008, Pisa, Italy*, pages 1109–1114, 2008.

[113] L. Yao, X. Wang, X. Wang, H. Hu, and G. Wu. Publishing sensitive trajectory data under enhanced l-diversity model. In *MDM*, pages 160–169. IEEE, 2019.

[114] R. Yarovoy, F. Bonchi, L. V. S. Lakshmanan, and W. H. Wang. Anonymizing moving objects: how to hide a MOB in a crowd? In *EDBT*, volume 360 of *ACM International Conference Proceeding Series*, pages 72–83. ACM, 2009.

[115] D. Yin and Q. Yang. Gans based density distribution privacy-preservation on mobility data. *Secur. Commun. Networks*, 2018:9203076:1–9203076:13, 2018.

[116] L. Yu, W. Zhang, J. Wang, and Y. Yu. Seqgan: Sequence generative adversarial nets with policy gradient. In *AAAI*, pages 2852–2858. AAAI Press, 2017.

[117] H. Zang and J. Bolot. Anonymization of location data does not work: a large-scale measurement study. In *MobiCom*, pages 145–156. ACM, 2011.

[118] Y. Zhan, A. Kyllo, A. Mashhadi, and H. Haddadi. Privacy-aware human mobility prediction via adversarial networks. *CoRR*, abs/2201.07519, 2022.

[119] J. Zhang, X. Xiao, and X. Xie. Privtree: A differentially private algorithm for hierarchical decompositions. In *SIGMOD Conference*, pages 155–170. ACM, 2016.

[120] X. Zhao, D. Pi, and J. Chen. Novel trajectory privacy-preserving method based on prefix tree using differential privacy. *Knowl. Based Syst.*, 198:105940, 2020.

[121] Y. Zheng. Trajectory data mining: An overview. *ACM TIST*, 6(3):29:1–29:41, 2015.

[122] B. Zhou and J. Pei. The *k*-anonymity and *l*-diversity approaches for privacy preservation in social networks against neighborhood attacks. *Knowl. Inf. Syst.*, 28(1):47–77, 2011.

[123] J. G. Zilly, R. K. Srivastava, J. Koutník, and J. Schmidhuber. Recurrent highway networks. In *ICML*, volume 70 of *Proceedings of Machine Learning Research*, pages 4189–4198. PMLR, 2017.