Rigorous and Flexible Privacy Models for Utilizing Personal Spatiotemporal Data

Yang Cao
Supervised by Masatoshi Yoshikawa
Kyoto University, Japan
soyo@db.soc.i.kyoto-u.ac.jp

ABSTRACT

Personal data are the new oil. Vast amounts of spatiotemporal data generated by individuals have been collected and analyzed, such as check-in data, trajectories, web browsing data and timestamped medical records. These personal data are a valuable resource for businesses and also have the potential to provide significant social benefits through sharing and reuse. However, privacy concerns hinder the wider use of these personal data. To this end, dozens of privacy protection models have been proposed for *privacy-preserving data publishing* (PPDP). ϵ-differential privacy has emerged as a de facto privacy model for PPDP because of its rigorous theoretical guarantees, but it has also been criticized as impractical in many scenarios. We attempt to design rigorous and flexible privacy models that able to bridge the gap between theoretical limitations and practical needs. In this article, we first motivate the importance of rigorousness and flexibility for privacy models and then present two privacy models that extend differential privacy in a practical manner.

1. INTRODUCTION

New opportunities are arising to enrich our understanding of the physical world by utilizing personal big data. Such data generated by individuals that possess spatial and temporal properties are called *personal spatiotemporal data* and include information such as people's movements, search logs, shopping records, social activities and medical histories. Companies such as Google, Amazon and Facebook have created enormous wealth by using such personal data. These data also have the potential to support significant initiatives for the public good, such as innovative services, traffic monitoring, disease surveillance and health promotion. Therefore, it is important to enable the sharing and reuse of personal spatiotemporal data.

However, several well-known cases of privacy leakage have arisen in attempts to provide privacy-preserving data publishing (PPDP) [1]. One of the most famous privacy leakage cases occurred with regard to the Massachusetts Group Insurance Commission (GIC) medical dataset, which contains anonymized medical records and was first made available in 1996 for medical research and public health purposes. Latanya Sweeney [2] linked the anonymized GIC database (which retained the birthday, sex, and ZIP code of each patient) with voter registration records to identify the medical records of the governor of Massachusetts. In 2006, AOL Inc. released more than 650 thousand users' anonymized search logs for research purposes, but researchers [3] found it is possible to de-anonymize some users by mining these timestamped records. A recent privacy leakage case was related to the datasets released for the Netflix Prize competition for recommender algorithms. Researchers [4] verified that the datasets could be de-anonymized by linking them with the IMDB dataset; as a result, Netflix became the target of a lawsuit and was obliged to terminate this well-known competition in 2010. In Japan, the East Japan Railway Company (JR) faced a scandal in 2013 because of the selling of customer data collected through the use of SUICA cards. These privacy leakage cases illustrate the conflict between increasing public concerns about data privacy and the need for personal data publishing. Therefore, privacy issues have become a critical problem in the big data era.

Personal spatiotemporal data are highly sensitive even if they are anonymized. Studies [5] [6] have shown that only four spatiotemporal data points from anonymized mobile datasets and credit card records (with purchase times and locations) are sufficient to uniquely identify 90% of individuals. Another study [7] has demonstrated that it is possible to track smart phone locations simply by monitoring battery consumption data (timestamps and remaining battery life), which do not require the users' permission to be shared. Therefore, because of the unique and rich patterns that are present in personal spatiotemporal data, a rigorous privacy model is needed to utilize personal spatiotemporal data in a privacy-preserving manner.

At the same time, privacy models should be flexible for the following two reasons. First, because privacy protection implies the hiding or perturbing of sensitive information[1], the privacy level should be tunable (to an appropriate level) for different applications to achieve a suitable trade-off between privacy and utility. Second, privacy as a complex social concept should be modifiable in accordance with different social contexts. For example, different users might have different

[1] Although some cryptographic approaches (e.g., homomorphic encryption) can be used for PPDP, because they limit the possible data recipients, we here discuss primarily perturbation-based approaches that enable a wider sharing of data.

Proceedings of the VLDB 2016 PhD Workshop, September 9, 2016. New Delhi, India.

preferences regarding privacy; for a single user, the desired protection levels might also be different at different times and places.

In the following, we briefly review the background for this research in Section 2 and then present two proposed privacy models in Section 3 and Section 4.

2. BACKGROUND

2.1 Previous Privacy Models

Previous studies have considered two major families of privacy models: k-anonymity [2] and ϵ-differential privacy [8]. k-anonymity [2] was proposed in 2002 and was the first privacy model to be widely accepted by practitioners. However, Machanavajjhala [9] identified a serious flaw of k-anonymity by demonstrating the possibility of homogeneity attacks and background knowledge attacks and, to resolve this flaw, proposed an improved privacy model based on k-anonymity named ℓ-diversity. However, ℓ-diversity had its own shortcomings, which motivated further improvement to yield subsequent new privacy models, such as t-closeness [10] and M-invariance [11]. Researchers have realized that the essential defect of the k-anonymity family is that the privacy guarantee depends on the background knowledge possessed by adversaries. In 2006, Dwork et al. proposed ϵ-differential privacy [8], which has received increasing attention because it is guaranteed to be rigorous and mathematically provable. Differential privacy ensures protection against adversaries with *arbitrary* background knowledge.

2.2 Differential Privacy

Intuitively, differential privacy guarantees that any single user's data in the database have only a "slight" (bounded by ϵ) impact on changes to the outputs. The parameter ϵ that is used to control the privacy level is defined as a real positive number. A lower value of ϵ corresponds to a lower likelihood of privacy leakage. In the case of a suitably small ϵ, an adversary cannot associate any piece of information with a specific individual by mining the answers to queries.

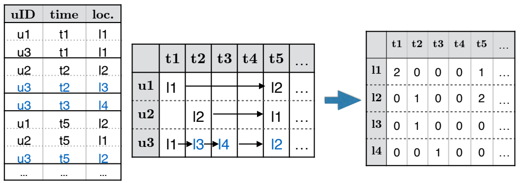
Definition 1 (Neighboring Database [8])**.** *If D is a database and D' is nearly identical to D but differs by one record, then we say that D and D' are two neighboring databases.*

Definition 2 (ϵ-Differential Privacy, ϵ-DP [8])**.** *Let Λ be a randomized algorithm, and let $\boldsymbol{R}$ represent all possible outputs of Λ. The algorithm Λ satisfies ϵ-DP if the following holds for any $r \in \boldsymbol{R}$ and any two neighboring databases D and D'.*

$$Pr[\Lambda(D) = r] \leq exp(\epsilon) \cdot Pr[\Lambda(D') = r] \tag{1}$$

In the inequality expressed in (1), a positive parameter ϵ, called the *privacy budget*, is given in advance. It is used to control the level of privacy protection. A lower values of ϵ corresponds to higher privacy and more randomness, and vice versa. The *global sensitivity* (or *sensitivity* for short) $GS(Q)$ of query Q is the maximum distance L_1 between the query results for any two neighboring databases.

Two widely used methods of achieving differential privacy are the Laplace mechanism [12], which adds random noise to the real data to prevent the disclosure of sensitive information, and the Exponential mechanism [13], which randomly returns the desired item with some calibrated probability.



uID	time	loc.
u1	t1	l1
u3	t1	l1
u2	t2	l2
u3	t2	l3
u3	t3	l4
u1	t5	l2
u2	t5	l1
u3	t5	l2
...	...	...

(a) Raw data.

	t1	t2	t3	t4	t5	...
u1	l1	→	→	→	l2	...
u2		l2	→	→	l1	...
u3	l1 →	l3 →	l4	→	l2	...

(b) Trajectory data.

	t1	t2	t3	t4	t5	...
l1	2	0	0	0	1	...
l2	0	1	0	0	2	...
l3	0	1	0	0	0	...
l4	0	0	1	0	0	...

(c) Aggregate data.

Figure 1: Raw data and aggregate information.

Below, we summarize two crucial weaknesses of differential privacy that hinder its extensive use in data publishing.

- **One size fits all.** ϵ-differential privacy uses a single parameter ϵ to represent the protection level for all users in a database rather than providing a personalized privacy protection level.

- **Data independence assumption.** Differential privacy assumes that the tuples in a database are independent; however, real-world data tend to be temporally or spatially correlated.

2.3 Problem setting

Here, we describe a scenario of the publishing of aggregate spatiotemporal data. We consider a scenario in which a trusted server collects spatiotemporal data points from users and continuously stores them in database D_i at each timestamp $i \in [1, t]$. Let t denote the current timestamp. Each data point $\langle uID, time, loc\rangle$ is a row in D_i (Fig. 1(a)). Suppose that $\boldsymbol{locs}$ is the set of all locations in which we are interested (POIs); then, the server wishes to publish a vector $\boldsymbol{r}_i$ at each timestamp i that consists of the counts of $loc \in \boldsymbol{locs}$ that appear in D_i (Fig.1(c)), i.e., the answer to the count query $Q^c : D_i \rightarrow \mathbb{R}^{|\boldsymbol{locs}|}$. Without loss of generality, we assume that each user appears at only *one* location at most at each timestamp. We need to transform the aggregate data depicted in Figure 1(c) into a secure form for publishing because they are computed directly from the sensitive raw data. The most straightforward method is to employ the Laplace mechanism [12] (add random Laplace noise) to achieve DP. In accordance with the limitations of DP discussed above, two problems are encountered when applying DP in practice, as follows:

- How can personalized differential privacy be achieved?
- How can differential privacy be guaranteed even if the data are temporally or spatially correlated?

We present two privacy models, *ℓ-trajectory privacy* (Section 3) and *spatiotemporal privacy* (Section 4), to solve the above problems.

3. ℓ-TRAJECTORY PRIVACY

In a previous work [14], we defined a new privacy model that we called *ℓ-trajectory privacy* to guarantee that any ℓ-length trajectories for each user are protected under differential privacy. We formalized our privacy definition, proved its feasibility, and designed efficient mechanisms to implement it. This privacy model is personalized, i.e., different users can specify different lengths of ℓ-trajectories (ℓ successive data points) depending on their privacy requirements, as illustrated in Figure 2(b). A closely related privacy model called w-event privacy [15] protects the data within sliding windows (all data points in w with successive timestamps), as shown in Figure 2(a). In the following, we present the definition of the ℓ-trajectory privacy model and related experimental results.

We first define the neighboring databases in our setting.

Definition 3 (ℓ-Trajectory Neighboring Stream Prefixes)**.** *Let S'_t be a near copy of the trajectory stream prefixes S_t that differs in ℓ neighboring databases D'_i. S_t and S'_t are neighboring ℓ-trajectory stream prefixes if one can be obtained from the other by modifying single or multiple locs in any one ℓ-trajectory $\ell_{u,k}$ (recall that an ℓ-trajectory is a set of ℓ spatiotemporal data points). We say that S_t and S'_t are neighboring with respect to $\ell_{u,k}$.*

Intuitively, ℓ-trajectory privacy attempts to ensure that any ℓ-trajectory for any single user has a only "slight" (bounded by ϵ) impact on the outputs.

Definition 4 (ℓ-Trajectory ϵ-Differential Privacy)**.** *Let Λ be an algorithm that takes prefixes of trajectory streams $S_t = \{D_1, \cdots, D_t\}$ as inputs. Let $\boldsymbol{N}_t = \{\boldsymbol{n}_1, \cdots, \boldsymbol{n}_t\}$ be a possible perturbed output stream of Λ. If the following holds for any ℓ-trajectory neighboring S_t and S'_t,*

$$Pr\left[\Lambda(S_t) = \boldsymbol{N}_t\right] \leq e^{\epsilon} \cdot Pr\left[\Lambda(S'_t) = \boldsymbol{N}_t\right] \qquad (2)$$

then we say that Λ satisfies ℓ-trajectory ϵ-differential privacy (or ℓ-trajectory privacy for brevity).

The following theorem provides insight regarding the implementation of ℓ-trajectory privacy.

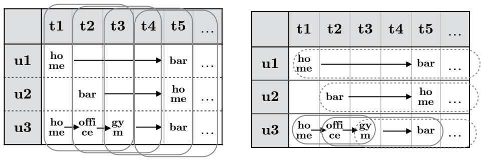
Theorem 1. *Let Λ be an integrated algorithm that takes stream prefixes $S_t = \{D_i, i \in [1, t]\}$ as inputs and produces $\boldsymbol{N_t} = \{\boldsymbol{n}_i, i \in [1, t]\}$ as outputs. Λ consists of a series of sub-algorithms $\{A_i, i \in [1, t]\}$, each of which takes D_i as its input and outputs noisy data $\boldsymbol{n}_i$ with independent randomness. Presume that A_i ensures ε_i-DP, and let $\tau_{u,k}$ be the set of timestamps dominated by an arbitrary trajectory $\ell_{u,k}$; then, Λ satisfies ℓ-trajectory privacy if*

$$\forall u, \forall k, \sum_{i \in \tau_{u,k}} \varepsilon_i \leq \epsilon \qquad (3)$$

Based on the above theorem, we designed an algorithmic framework for publishing differentially private aggregates that satisfy ℓ-trajectory privacy. The experimental results show that our approach is effective and efficient compared with previous works [16] and [15]. More details can be found in our paper [14].

4. $\epsilon_{\mathcal{ST}}$-DIFFERENTIAL PRIVACY

In existing studies, traditional DP mechanisms (e.g., the Laplace mechanism) are employed as primitives. These mechanisms assume that the data are independent or that adversaries do not have knowledge of the data correlations. However, data collected from the real world tend to be temporally or spatially correlated, and such correlations could be acquired by adversaries. In this study, we introduce a new class of adversaries named *realistic adversaries* who have a varying degree of knowledge of the spatiotemporal correlations of the data of interest. We demonstrate that the risk of a DP mechanism may increase over time given the existence of such adversaries. To prevent such privacy leakage, we propose a new privacy definition, *$\epsilon_{\mathcal{ST}}$-differential privacy*, and present novel mechanisms to satisfy this privacy definition. In the following, we first demonstrate the problem by means of a motivating example, present the definition of realistic adversaries, and briefly describe measures for protecting against such adversaries. This work has been submitted to an international conference.



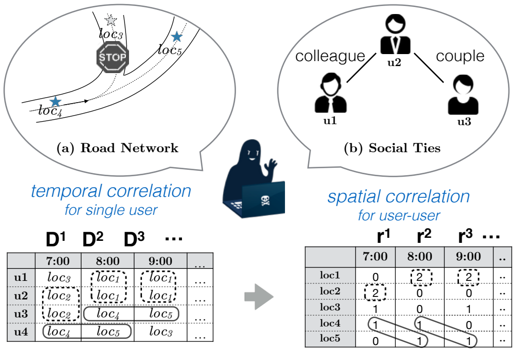
	t1	t2	t3	t4	t5	…
u1	home	→	→	→	bar	…
u2		bar	→	→	home	…
u3	home →	office →	gym	→	bar	…

(a) w-event privacy (w=3).

	t1	t2	t3	t4	t5	…
u1	home	→	→	→	bar	…
u2		bar	→	→	home	…
u3	home →	office →	gym	→	bar	…

(b) l-trajectory privacy (l=3).

Figure 2: Illustration of ℓ-trajectory privacy compared with w-event privacy [15].



(a) Road Network

(b) Social Ties

	7:00	8:00	9:00	…
u1	loc_3	loc_1	loc_1	…
u2	loc_2	loc_1	loc_1	…
u3	loc_2	loc_4	loc_5	…
u4	loc_4	loc_5	loc_3	…

(c) Spatiotemporal Data

	7:00	8:00	9:00	..
loc1	0	2	2	..
loc2	2	0	0	..
loc3	1	0	1	..
loc4	1	1	0	..
loc5	0	1	1	..

(d) Raw Aggregates

Figure 3: The vulnerability of differential privacy to correlated spatiotemporal data.

4.1 A Motivating Example

The following example (illustrated in Figure 3) shows that the existence of adversaries with knowledge of the spatiotemporal correlations of the data may degrade the expected privacy guarantee of DP.

Example 1. *Consider the scenario of spatiotemporal data publishing illustrated in Figure 3. A trusted server continuously collects users' locations at each time point. Our goal is to publish the aggregates shown in Table (d) while protecting against adversaries who may have some knowledge regarding membership in the database. We assume that each user appears at only one location at most at each time point. Then, according to the Laplace mechanism [12], adding $Lap(1/\epsilon)$ noise*[2] *to perturb the data can achieve ϵ-DP. However, from auxiliary information such as road network constraints, an attacker may know the mobility patterns of a targeted victim; for example, the adversary may know that the victim "always arrives at loc_5 after visiting loc_4," as illustrated in Figure 3(a), which leads to the patterns indicated by the solid lines in Figure 3(c) and (d). Consequently, because of the composability of DP, adding $Lap(1/\epsilon)$ noise guarantees only 2ϵ-DP against this attacker. We refer to the transition pattern of a single user as temporal correlation. By contrast, when an attacker has the following knowledge (illustrated in Figure 3(b)), privacy leakage will also occur: (1) Users u_1 and u_2 are always at the same location during working hours (e.g., they are colleagues). (2) Users u_2 and u_3 are always at the same location during leisure time (e.g., they are newlyweds). We refer to such geographical similarities between users as spatial correlations. In Figure 3(c) and (d), it is easy to see the resulting patterns, which are illustrated as dashed lines. Therefore, because of the high sensitivity induced by the highly correlated nature of the data, the addition of $Lap(1/\epsilon)$ noise achieves only 2ϵ-DP.*

[2] $Lap(b)$ denotes a Laplace distribution with variance $2b^2$.

4.2 Realistic Adversaries

To formalize realistic adversaries (RAs) and their impact on privacy leakage, we first study commonly used models for describing temporal and spatial correlations as means of expressing the prior knowledge of RAs. We then define a new privacy model, $\epsilon_{\mathcal{ST}}$-Differential Privacy ($\epsilon_{\mathcal{ST}}$-DP), to account for RAs.

Definition 5 (Temporal Correlations)**.** *The temporal correlation of user i is described by a transition matrix $P_i \in \mathbb{R}^{|\boldsymbol{loc}|\times|\boldsymbol{loc}|}$, which represents $\Pr(l_i^{t-1}|l_i^t)$, where l_i^t is the location of user i at time t.*

Definition 6 (Spatial Correlations)**.** *Let x_i denote Gaussian random variables that represent the possible locations of each user $i \in [1, n]$. GMRF $\mathbb{G}$, which we call the correlation graph, describes the spatial correlations among $x_1, \dots, x_n$.*

We define realistic adversaries in a comprehensive and flexible manner such that various scenarios can be represented.

Definition 7 (Realistic Adversaries)**.** *The adversaries targeting user i who possess various levels of prior knowledge, denoted by $R_i(P_i, G_i, m)$, are called the realistic adversaries of that user. There are three basic types of realistic adversaries: (1) $R_i(P_i, \emptyset, m)$, (2) $R_i(\emptyset, G_i, m)$, and (3) $R_i(P_i, G_i, m)$.*

Attack by realistic adversaries is Bayesian in nature [17]. Let $D^t_{\mathcal{K}} \subset D^t$ represent the prior knowledge of such adversaries, and let $D^t_{\mathcal{U}}$ be the unknown tuples, i.e., $D^t = D^t_{\mathcal{K}} \cup D^t_{\mathcal{U}} \cup \{l_i^t\}$. The adversaries attempt to infer the location of user i at time t. To this end, they first infer $D^t_{\mathcal{U}}$ based on $D^t_{\mathcal{K}}$ and a guessed value of l_i^t (or $l_i^{t\prime}$, to represent a different guess) and then attempt to distinguish the difference between the two neighboring databases D^t and $D^{t\prime}$, where $D^{t\prime} = D^t_{\mathcal{K}} \cup D^t_{\mathcal{U}} \cup \{l_i^{t\prime}\}$.

Definition 8 (ϵ-DP$_{\mathcal{ST}}$)**.** *At each time t, a DP mechanism $\mathcal{M}^t$ takes D^t as input and produces $\boldsymbol{r}^t$ as output. The privacy leakage of $\mathcal{M}^t$ against $R_i(P_i, G_i, m)$ is defined as follows.*

$$PL(R_i, \mathcal{M}^t) \overset{\text{def}}{=} \sup_{\forall |D^t_{\mathcal{K}}|=m} \log \frac{\Pr(\boldsymbol{r}^1, \dots, \boldsymbol{r}^t | l_i^t, D^t_{\mathcal{K}})}{\Pr(\boldsymbol{r}^1, \dots, \boldsymbol{r}^t | l_i^{t\prime}, D^t_{\mathcal{K}})}$$

$$= \sup_{\forall |D^t_{\mathcal{K}}|=m} \log \frac{\sum_{D^t_{\mathcal{U}}} \Pr(\boldsymbol{r}^1, \dots, \boldsymbol{r}^t | D^t) \Pr(D^t_{\mathcal{U}} | l_i^t, D^t_{\mathcal{K}})}{\sum_{D^t_{\mathcal{U}}} \Pr(\boldsymbol{r}^1, \dots, \boldsymbol{r}^t | D^{t\prime}) \Pr(D^t_{\mathcal{U}} | l_i^{t\prime}, D^t_{\mathcal{K}})} \qquad (4)$$

$\mathcal{M}^t$ satisfies ϵ-DP$_{\mathcal{ST}}$ (i.e., DP under spatiotemporal correlations) if, for all $R_i, i \in [1, n]$, it holds that $PL(R_i, \mathcal{M}^t) \leq \epsilon$.

4.3 Quantifying Privacy Leakage

The primary challenge of protecting against realistic adversaries is to finely quantify the privacy leakage, i.e., Equation (4). Because of space limitation, we briefly describe a concept for how this can be achieved and present a preliminary result. We analytically divide Equation (4) into two parts, namely, the privacy leakages w.r.t. adversaries of type (1) (in Definition 7) and those of type (2). The corresponding problems of quantification can be translated into inference on the GMRF (for type (1)) and a linear-factional programming problem (for type (2)). We then analyze the combination of these problems, i.e., the privacy leakage w.r.t. type (3).

The result reveals that the privacy leakage of a DP mechanism w.r.t. realistic adversaries may *increase over time*, as shown in Example 2.

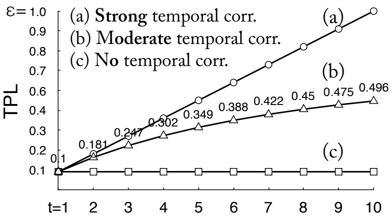


Figure 4: Privacy leakage may increase over time.

Example 2. *According to the Laplace mechanism, traditionally, adding $Lap(1/0.1)$ noise to published counts can achieve 0.1-DP at each time point. However, in an extreme case, if a user's data at any two successive time points are correlated with probability 1, say, $P_a = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, then continual data publishing is equivalent to releasing the* same *data multiple times. Hence, the privacy leakage at each time point will accumulate with respect to previous time points, exhibiting a linear increase (Figure 4(a)). In another extreme case, if there is no temporal correlation, or a uniform correlation such as $P_c = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, then the privacy leakage at each time point does not increase, as shown in Figure 4(c). Figure 4(b) depicts the privacy leakage induced by $P_b = \left(\begin{smallmatrix} 0.8 & 0.2 \\ 0 & 1 \end{smallmatrix}\right)$, which can be finely calculated using our algorithm.*

5. CONCLUSION AND FUTURE WORK

In this research, we addressed two fundamental limitations of differential privacy by developing *rigorous and flexible privacy models* for the use of personal spatiotemporal data. An interesting future direction of research will be to extend our privacy models to other data types.

Acknowledgments

We appreciate the comments of the anonymous reviewers. This research is partially supported by the Center of Innovation Program of the Japan Science and Technology Agency, JST.

6. REFERENCES

[1] Benjamin C. M. Fung, Ke Wang, Rui Chen, and Philip S. Yu. Privacy-preserving data publishing: A survey of recent developments. *ACM Comput. Surv.*, 42(4):14:1–14:53, 2010.

[2] Latanya Sweeney. K-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557–570, 2002.

[3] Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. Wherefore art thou r3579x?: Anonymized social networks, hidden patterns, and structural steganography. In *WWW*, pages 181–190, 2007.

[4] A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *S&P '08*, pages 111–125, 2008.

[5] Yves-Alexandre de Montjoye, CÃľsar A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. Unique in the crowd: The privacy bounds of human mobility. *Sci. Rep.*, 3, 2013.

[6] Yves-Alexandre de Montjoye, Laura Radaelli, Vivek Kumar Singh, and Alex âĂIJSandyâĂİ Pentland. Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, 347(6221):536–539, 2015.

[7] Yan Michalevsky, Gabi Nakibly, Gunaa Arumugam Veerapandian, Dan Boneh, and Gabi Nakibly. PowerSpy: location tracking using mobile device power analysis. In *SEC*, pages 785–800, 2015.

[8] Cynthia Dwork. Differential privacy. In *ICALP*, pages 1–12, 2006.

[9] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. L-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data*, 1(1):3–es, 2007.

[10] Ninghui Li, Tiancheng Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *IEEE 23rd International Conference on Data Engineering, 2007. ICDE 2007*, pages 106–115, 2007.

[11] Xiaokui Xiao and Yufei Tao. M-invariance: Towards privacy preserving re-publication of dynamic datasets. In *SIGMOD*, pages 689–700, 2007.

[12] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, pages 265–284, 2006.

[13] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *FOCS*, pages 94–103, 2007.

[14] Yang Cao and Masatoshi Yoshikawa. Differentially private real-time data publishing over infinite trajectory streams. *IEICE Trans. Inf.& Syst.*, E99-D(1):163–175, 2016.

[15] Georgios Kellaris, Stavros Papadopoulos, Xiaokui Xiao, and Dimitris Papadias. Differentially private event sequences over infinite streams. *PVLDB*, 7(12):1155–1166, 2014.

[16] Liyue Fan, Li Xiong, and Vaidy Sunderam. FAST: differentially private real-time aggregate monitor with filtering and adaptive sampling. In *SIGMOD*, pages 1065–1068, 2013.

[17] Daniel Kifer and Ashwin Machanavajjhala. Pufferfish: A framework for mathematical privacy definitions. *ACM Trans. Database Syst.*, 39(1):3:1–3:36, 2014.