Local Differential Privacy on Metric Spaces: Optimizing the Trade-Off with Utility

Mário S. Alvim, Konstantinos Chatzikokolakis, Catuscia Palamidessi, Anna Pazii

Metric-based local differential privacy for statistical applications

Mário S. Alvim*, Konstantinos Chatzikokolakis[†‡], Catuscia Palamidessi[§‡] and Anna Pazii[‡§]
*UFMG, Belo Horizonte, Brazil
[†]CNRS, France
[‡]Ecole Polytechnique, University of Paris Saclay, France
[§]INRIA, University of Paris Saclay, France

***Abstract*—Local differential privacy (LPD) is a distributed variant of differential privacy (DP) in which the obfuscation of the sensitive information is done at the level of the individual records, and in general it is used to sanitize data that are collected for statistical purposes. LPD has the advantage it does not need to assume a trusted third party. On the other hand LDP in general requires more noise than DP to achieve the same level of protection, with negative consequences on the utility. In practice, utility becomes acceptable only on very large collections of data, and this is the reason why LDP is especially successful among big companies such as Apple and Google, which can count on a huge number of users. In this paper, we propose a variant of LDP suitable for metric spaces, such as location data or energy consumption data, and we show that it provides a much better utility for the same level of privacy.**

***Keywords*-Local differential privacy, d_x-privacy, Kantorovich lifting.**

I. Introduction

With the ever-increasing use of internet-connected devices, such as computers, IoT appliances (smart meters, home monitoring devices), and GPS-enabled equipments (mobile phones, in-car navigation systems), personal data are collected in larger and larger amounts, and then stored and manipulated for the most diverse purposes. For example, the web browsing history can be used for profiling the user and sending him targeted publicity. Power-consumption data from smart meters can be analyzed to extract typical daily consumption patterns in households [1], or to identify the right customers to target for demand response programs [2]. Location data can be used to find the most frequented public areas (for instance, to deploy hotspots) [3], or to provide traffic information [4].

Undeniably, the Big Data technology provides enormous benefits to individuals and society, ranging from helping the scientific progress to improving the quality of service. On the other hand, however, the collection and manipulation of personal data raises alarming privacy issues. Not only the experts, but also the population at large are becoming increasingly aware of the risks, due to the repeated cases of violations and leaks that keep appearing on the news. It is particularly disturbing when personal data are collected without the user's consent, or even awareness. For instance, in 2011 it was discovered that the iPhone was storing and collecting location data about the user, syncing them with iTunes and transmitting them to Apple, all without the user's knowledge [5]. More recently, the Guardian has revealed, on the basis of the documents provided by Edward Snowden, that the NSA and the GCHQ have been using certain smartphone apps, such as the wildly popular Angry Birds game, to collect users' private information such as age, gender and location [6], again without the users' knowledge.

Another related problem is that users often do not have the possibility to control the precision and the amount of personal information that is being exposed. For instance the Tinder application was found sharing the exact latitude and longitude co-ordinates of users as well as their birth dates and Facebook IDs [7], and even after the initial problem was fixed, it was still sharing more accurate location data than intended, as users could be located to within 100 feet of their present location [8]. The leakage of *precise* personal information is particularly problematic, especially when considering that various kinds of personal data from different sources can be linked and aggregated into a user profile [9], [10], and can fall in the hands of malicious parties.

Until recently, the most popular and used data sanitization technique was anonymization (removal of names) or slightly more sophisticated variants like k-anonymity [11] ensuring indistinguishability within groups of at least k people, and ℓ-diversity, ensuring a variety of values for the sensitive data within the same group [12]. Unfortunately, these techniques have been proved unable to provide an acceptable level of protection, as several works have shown that individuals in anonymized datasets can be re-identified with high accuracy, and their personal information exposed (see for istance [13], [14]).

In the meanwhile, *differential privacy* (DP), has emerged and imposed itself as a convincing alternative to anonymity. Together with its distributed version *local differential privacy*, it represents the cutting-edge of research on privacy protection.

DP was developed in the area of statistical databases, and it aims at protecting the individuals' data while allowing to make public the aggregate information [15]. This is obtained by adding controlled noise to the query outcome, in such a way that the data of a single individual will have a negligible

impact on the reported answer. More precisely, let $\mathcal{M}$ be a (noisy) mechanism for answering a certain query on a generic database D, and let $P[\mathcal{M}(D) \in S]$ denote the probability that the answer given by $\mathcal{M}$ to the query, on D, is in the (measurable) set of values S. We say that $\mathcal{M}$ is ε-differentially private if for every pairs of *adjacent* databases D and D' (i.e., differing only for the value of a single individual), and for every measurable set S, we have $P[\mathcal{M}(D) \in S] \leq e^{\varepsilon} \cdot P[\mathcal{M}(D') \in S]$. DP has two important advantages with respect to other approaches: (a) it is independent from the side-information of the adversary, thus a differentially-private mechanism can be designed without taking into account the context in which it will have to operate, and (b) it is compositional, i.e., if we combine the information that we obtain by querying two differentially-private mechanisms, the resulting mechanism is also differentially-private. Furthermore, (c) differentially-private mechanisms usually provide a good trade-off between utility and privacy, i.e., they preserve the privacy of the individuals without destroying the utility of the collective data.

Local differential privacy (LDP) is a distributed variant of differential privacy in which users obfuscate their personal data by themselves, before sending them to the data collector [16]. Technically, an obfuscation mechanism $\mathcal{M}$ is locally differentially private with privacy level ε if for every pair of input values $x, x' \in \mathcal{X}$ (the set of possible values for the data of a generic user), and for every measurable set S, we have $P[\mathcal{M}(x) \in S] \leq e^{\varepsilon} \cdot P[\mathcal{M}(x') \in S]$. The idea is that the user provides $\mathcal{M}(x)$ to the data collector, and not x. In this way, the data collector can only gather, stock and analyze the obfuscated data. Based on these he can infer statistics (e.g., histograms, or heavy hitters [17]) of the original data. LPD implies DP on the collected data, and has the same advantages of independence from the side-information and compositionality. Furthermore, with respect to the centralized model, it has the further advantages that (a) each user can choose the level of privacy he wishes, (b) it does not need to assume a trusted third party, and (c) since all stored records are individually-sanitized , there is no risk of privacy breaches due to malicious attacks. LDP is having a considerable impact, specially after large companies such as Apple and Google have started to adopt it for collecting the data of their customers for statistical purposes [18].

The disadvantage of LDP is that it can spoil substantially the utility of the data. Even in those cases where the trade-off with utility is most favorable, namely the statistical applications, it is usually necessary to have a huge collection of data in order for the statistics to be significant. Fortunately, however, the data domains are often equipped with structures that could be exploited to improve utility. In these notes, we focus on data domains that are provided with a notion of distance. This is the case, for instance, of location data, energy consumption in smart meters, age and weight in medical records, etc. Usually, when these data are collected for statistical purposes, the accuracy of the distribution is measured also with respect to the same notion of distance. In such scenarios, *we argue that the trade-off between privacy and utility can be greatly improved by exploiting the concept of approximation intrinsic in metrics*.

Following this intuition, we propose a variant of local differential privacy based on the notion of $d_{\mathcal{X}}$-privacy intoduced in [19]. An obfuscation mechanism $\mathcal{M}$ is $\varepsilon \cdot d_{\mathcal{X}}$-private if for every pair of input values $x, x' \in \mathcal{X}$, and for every measurable set S, we have $P[\mathcal{M}(x) \in S] \leq e^{\varepsilon \cdot d_X(x,x')} \cdot P[\mathcal{M}(x') \in S]$. In other words, $d_{\mathcal{X}}$-privacy relaxes the privacy requirement by allowing two data to become more and more distinguishable as their distance increases. Thus, it allows the adversary to infer some approximate information about the true value, but it does not allow him to infer the exact true value. As explained in [19], $d_{\mathcal{X}}$-privacy can be implemented by using an extended notion of Laplacian noise, or of geometric noise.

The original motivation for the notion of $d_{\mathcal{X}}$-privacy was for real-time punctual applications. In particular, the instance of $d_{\mathcal{X}}$-privacy in which $d_{\mathcal{X}}$ is the geographical distance has been used in the context of location privacy, under the name of *geo-indistinguishability*, to protect the location of the user during the interaction with location-based services (LBSs) [20], [21]. These are services that provide the user with certain desired information which depends on the location communicated by the user, like for instance points of interest (POI) near the location. The idea is that the user does not need to communicate his exact coordinates, an approximate location should suffice to obtain the requested information without too much degradation of the quality of service.

Geo-indistinguishability has been quite successful, and its implementation via the Laplacian mechanism has been adopted as the basis or as a component of several tools and frameworks for location privacy, including: LP-Guardian [22], LP-Doctor [23], STAC [24], Location Guard [25], and the SpatialVision QGIS plugin [26]. Here, we want to show that geo-indistinguishability, and more in general $d_{\mathcal{X}}$-privacy, can also be used to protect privacy when collecting data for statistical purposes, and that if the statistics are distance-sensitive, then $d_{\mathcal{X}}$-privacy preserves the utility of the data better than the standard LDP methods.

In the rest of these notes, we will discuss the improvement on trade-off utility-privacy compared to standard LDP methods, and show some experimental results based on the Gowalla dataset [27], [28]. For simplicity we will restrict the analysis to the case of discrete metric spaces. We will consider, in particular, the mechanisms of K-ary Randomized Response (K-RR) [29] for LPD, and the (discretized) Laplacian and geometric mechanisms for $d_{\mathcal{X}}$-privacy.

II. Utility

We consider a notion of utility suitable for statistical applications. The scenario is the following: let $\mathcal{X}$ (the universe) be a set of secrets, endowed with a notion of distance $d_{\mathcal{X}}$, and let $\mathbb{D}\mathcal{X}$ be the set of distributions on $\mathcal{X}$. Let D be an unsanitized dataset on $\mathcal{X}$, namely a multiset of elements of $\mathcal{X}$ (i.e., an histogram), determining a distribution $\pi \in \mathbb{D}\mathcal{X}$. Assume that each individual element x in D gets sanitized by injection of noise, thus producing a noisy dataset $\hat{D}$. From $\hat{D}$ we then try to reconstruct the distribution π as well as we can, assuming that we only know $\hat{D}$ and the mechanism $\mathcal{M}$ for noise injection.

In order to reconstruct as precisely as possible the original π, we propose to use the Expectation-Maximization (EM) method [30], also known as Iterative Bayesian Update, which iteratively estimates the distribution until convergence to a fixed point. The feature of this method is that the final estimate (converged value) is equal to the Maximum Likelihood estimate in the probability simplex, and it is shown in [30] that it significantly outperforms the other known techniques like the matrix inversion method.

Let $\hat{\pi} \in \mathbb{D}\mathcal{X}$ be the output of the EM method. Intuitively, the *utility loss* with respect to the original database should reflect the *expected difference between the statistical properties based on the noisy data and those based on the real data*. To formalize the notion of *expectation*, we can regard $\mathcal{M}$, in abstract terms, as a device that inputs π and output a set of possible distributions $\hat{\pi}_1, \hat{\pi}_2, \ldots \hat{\pi}_i \ldots$, each with a certain probability $p_1, p_2, \ldots p_i \ldots$. In other words, $\mathcal{M}$ can be seen as a transformation that associates to each π a function Δ which assigns a probability mass to every distribution, i.e, $\Delta(\hat{\pi}_i) = p_i$. This type of functions Δ are called *hyperdistributions* in [31], [32]. Note that also π can be seen as a hyperdistribution: it is the function that assigns 1 to π, and we will denote it by $[\pi]$.

Concerning the *difference between statistical properties*: in very general terms, we can represent a statistical property as a functions $f : \mathbb{D}\mathcal{X} \to \mathbb{R}$, where $\mathbb{R}$ is the set of reals. We want to capture as many f's as possible, but it is reasonable to assume that the difference between $f(\pi)$ and $f(\pi')$ must be bound by the distance between the distributions π and π', for some "reasonable" notion of distance $d_{\mathbb{D}\mathcal{X}}$. In other words, we want to avoid that negligible differences on the distributions may produce unbound differences in the statistics. For this reason, we restrict the statistics of interest, $\mathcal{F} \subseteq (\mathbb{D}\mathcal{X} \to \mathbb{R})$, to be the set of 1-Lipshitz[1] functions with respect to $d_{\mathbb{D}\mathcal{X}}$. Hence, $\mathcal{F} = \{f : \mathbb{D}\mathcal{X} \to \mathbb{R} \mid f \text{ is 1-Lipshitz w.r.t. } d_{\mathbb{D}\mathcal{X}}\}$. Finally, since we want to abstract from the peculiarity of any particular statistics, we will consider the *maximum difference* induced by the noise on all the statistics in $\mathcal{F}$. Summarizing, we can define the utility loss as:

$$\mathcal{UL}(\mathcal{M}, \pi, d_{\mathbb{D}\mathcal{X}}) \;=\; \max_{f \in \mathcal{F}} \mid \sum_{\hat{\pi}} \Delta(\hat{\pi}) f(\hat{\pi}) - f(\pi) \mid \qquad (1)$$

where $\Delta = \mathcal{M}(\pi)$. It is worth noting that the rhs of this definition is the distance between Δ and $[\pi]$ obtained as the *Kantorovich lifting* of $d_{\mathbb{D}\mathcal{X}}$, which we will denote as $K(d_{\mathbb{D}\mathcal{X}})(\Delta, [\pi])$. Following the same intuition, we can define the distance $d_{\mathbb{D}\mathcal{X}}$ as the *Kantorovich lifting* of $d_{\mathcal{X}}$, thus establishing a link with the ground distance on the domain of secrets.

III. Tuning privacy

The notions of LDP and of $d_{\mathcal{X}}$-privacy both depend on privacy parameters ε's, but these ε's do not represent the same level of privacy in the two definitions. They are not even of the same type: the ε in LDP is a pure number, while the ε in $d_{\mathcal{X}}$-privacy is the converse of a distance. Therefore, in order to compare the utility of an LDP mechanism $\mathcal{M}$ with that of a $d_{\mathcal{X}}$-private mechanism $\mathcal{M}'$, we have first to tune their privacy parameters so to ensure that $\mathcal{M}$ and $\mathcal{M}'$ provide the same privacy protection. To this end, we consider the notion of location privacy proposed in [33], defined as expected distance between the reported location and the real location. [2] Generalizing to $d_{\mathcal{X}}$-privacy, we require that $\mathcal{M}$ and $\mathcal{M}'$ give the same expected distance between $x \in D$ and the corresponding reported datum. Namely:

$$\begin{gathered} \textstyle\sum_{x,y \in \mathcal{X}} \pi(x) P[\mathcal{M}(x) = y] d_{\mathcal{X}}(x, y) \\ = \\ Ed \\ = \\ \textstyle\sum_{x,y \in \mathcal{X}} \pi(x) P[\mathcal{M}'(x) = y] d_{\mathcal{X}}(x, y) \end{gathered} \qquad (2)$$

where Ed represents the desired level of protection, expressed in terms of the expected distance of the reported location from the real one.

IV. The mechanisms

We now recall the definition of the K-ary RR mechanism [29], representative of LPD, and the Laplacian and geometric mechanisms, representative of $d_{\mathcal{X}}$-privacy.

A. The Laplacian mechanism

The Laplacian mechanism $\mathcal{M}_L$ is used when $(\mathcal{X}, d_{\mathcal{X}})$ is a continuous metric space. Given a real location x, it reports a location y with a probability density function (pdf) defined as:

$$dP_x(y) = \lambda_L \, e^{-\varepsilon \cdot d_{\mathcal{X}}(x,y)}$$

[1] The requirement of 1-Lipshitz is not really essential, it could be k-Lipshitz for an arbitrary k. The important constraint is that the difference on the statistics is bound in some uniform way by the difference on the distributions.

[2] The definition in [33] also takes into account the knowledge of the prior, and the possibility to remap the reported location in the most most likely one according to the additional information provided by the prior. In our case, we want a definition that does not depend on the knowledge of the prior (since π is supposed to be unknown), and without the prior information, for the mechanisms we consider, the most likely location is always the reported one. Hence we do not need remapping.

where λ_L is a normalization factor.

In case we want to work in a discrete setting, we can first discretize the metric space by partitioning $\mathcal{X}$ into cells and defining the distance between two cells as the $d_{\mathcal{X}}$ between the centers of the cells. Then we can discretize the mechanism by defining the probability of a cell C as the probability mass obtained by the integration of the pdf over C.

B. The geometric mechanism

The geometric mechanism $\mathcal{M}_G$ is used when $(\mathcal{X}, d_{\mathcal{X}})$ is a discrete metric space. It is defined similarly to Laplacian mechanism, with the exception that now x and y represent discrete locations, and we have a (discrete) probability distribution rather than a pdf:

$$P[\mathcal{M}_G(x) = y] = \lambda_G \, e^{-\varepsilon \cdot d_{\mathcal{X}}(x,y)}$$

where λ_G is a normalization factor.

If $(\mathcal{X}, d_{\mathcal{X}})$ is the result of a discretization of a continuous metric space, then the discretized laplacian is very similar, but not identical to the geometric mechanism.

V. The K-RR mechanism

The K-RR mechanism, aka *flat* mechanism, $\mathcal{M}_F$ is one of the simplest LPD mechanisms. The idea is that the result of the sanitization is a bit more likely to be the true value x than any other value in the domain (taken individually), and that on the other values the probability is distributed uniformly :

$$P[\mathcal{M}_F(x) = y] = \begin{cases} \frac{e^{\epsilon}}{|X|-1+e^{\epsilon}} & \text{if } y = x \\ \frac{1}{|X|-1+e^{\epsilon}} & \text{if } y \neq x \end{cases}$$

In [29], the k-RR mechanism has been shown to be optimal in the low privacy regime for a large class of information theoretic utility functions.

VI. Experimental results

In this section we compare the utility of the privacy mechanisms $\mathcal{M}_F$, $\mathcal{M}_G$ and the (discretized) $\mathcal{M}_L$ introduced in previous section, using a distribution derived from the Gowalla dataset, which contains location data (check-ins) relative to a certain population of users.

We consider an area of 4.5 Km $\times$ 4.5 Km in Paris, centered in 5 Boulevard de Sébastopol, near Le Halles. We discretize that area by considering a grid of 30×30 cells, so that every cell is 150 m $\times 150$ m, see Figure 1. These cells represent the elements of $\mathcal{X}$, and the distance $d_{\mathcal{X}}$ is defined as the geographic distance between the centers of the cells.

We consider 750 check-ins from Gowalla in this area, selected randomly, and we consider the multiset D obtained by counting the number of check-ins in each cell. Let π be the corresponding distribution.

We now tune the privacy parameters of $\mathcal{M}_F$, $\mathcal{M}_G$ and $\mathcal{M}_L$ so that the expected distance Ed of the reported location from the real one is the same for all of them (cfr. Requirement (2)). We set Ed to be 3 times the size of a cell, namely 450 m. The values of ε that we derive are: 8.24797 for $\mathcal{M}_F$, 0.00398441 for $\mathcal{M}_G$, and 0.00404249 for $\mathcal{M}_L$.

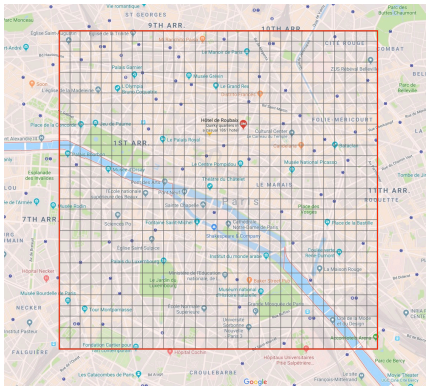

Figure 1. The area of Paris consider for evaluating the utility of the three mechanisms.

In order to compute the utility loss for these three mechanisms, we use the well-known fact that the Kantorovich distance is equal to the earth mover's distance (EMD) also known as the Wasserstein metric. Namely, we can equivalently rewrite (1) as

$$\mathcal{UL}(\mathcal{M}, \pi, d_{\mathbb{D}\mathcal{X}}) = \min_{\alpha} \sum_{\hat{\pi} \in dom(\Delta)} \alpha(\hat{\pi}, \pi) d_{\mathbb{D}\mathcal{X}}(\hat{\pi}, \pi) \quad (3)$$

where $\Delta = \mathcal{M}$, α ranges on the couplings that have as marginals Δ and $[\pi]$, and $d_{\mathbb{D}\mathcal{X}} = K(d_{\mathcal{X}})$. In conclusion, to determine the utility loss we need to compute the expectation of the Kantorovich distance between the reported location and the real location. We have done it for an increasing sequence of dataset $\emptyset \subseteq D_1 \subseteq D_2 \subseteq \ldots \subseteq D_n \subseteq \ldots \subseteq D$ constructed incrementally by adding each time 10 elements from the 750 check-ins, selected randomly, until all of them are inserted. The results are reported in Figure 2.

As we can see from this figure, the geometric and the discretized Laplacian have similar utility, and perform much better (in terms of utility) than the flat mechanism. In fact, if we consider statistics that are somehow coherent with the ground distance $d_{\mathcal{X}}$, then the fact that the geometric and the discretized Laplacian tend to assign a negligible probability to locations that are far away from the real one means that those locations do not significantly contribute to the loss of utility. In contrast, the flat mechanism treats in the same way all locations, independently from their distance from the real one. Consequently there are several locations that are far away and still carry a significant probability mass, thus taking a heavy toll on the utility.

Furthermore, we can see that the utility loss of the Flat mechanisms, although showing a tendency to diminish as the numbers of check-ins increases, it does so very slowly.

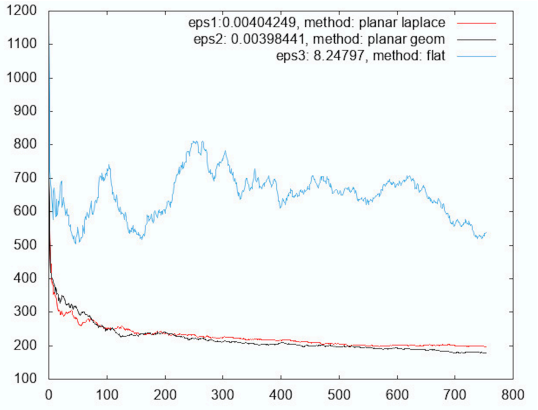

Figure 2. The utility loss for an increasing sequence of datasets taken from the Gowalla check-ins in the area illustrated in Figure 1. The numbers in the horizontal axis represent the number of check-ins $\times$ 10. The numbers in the vertical axis represent the distance, expressed in meters.

Finally, at the beginning (for less than 2000 check-ins) the behavior of the flat mechanism is extremely unstable. This is due again to the fact that the reported locations tend to be scattered in the whole area with high probability, which determine high fluctuations especially at the beginning when the data are few, as the addition of new data can cause a big change in the distributions.

VII. Conclusion

We have advocated the use of $d_{\mathcal{X}}$-privacy to protect privacy when data are collected for statistical purposes on domains of secrets endowed with a notion of distance, arguing that in such context $d_{\mathcal{X}}$-privacy offers a better trade-off between privacy and utility than traditional LPD methods. We have confirmed this claim by performing experimental evaluations of the utility of $d_{\mathcal{X}}$-private mechanisms and LPD ones on real location data from the Gowalla dataset. The results show that the gap in terms of utility (for the same level of privacy) is actually quite significant.

Acknowledgment

This work has been supported by the ANR project REPAS and by the project Epistemic Interactive Concurrency (EPIC) from the 862 STIC AmSud Program. Maário S. Alvim was supported by CNPq, CAPES, and FAPEMIG.

References

[1] H. Hino, H. Shen, N. Murata, S. Wakao, and Y. Hayashi, "A versatile clustering method for electricity consumption pattern analysis in households," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 1048–1057, 2013.

[2] C. Chelmis, J. Kolte, and V. K. Prasanna, "Big data analytics for demand response: Clustering over space and time," in *Proceedings of the IEEE International Conference on Big Data (Big Data)*, Oct 2015, pp. 2223–2232.

[3] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from gps trajectories," in *Proceedings of the 18th International Conference on World Wide Web*, ser. WWW '09. ACM, 2009, pp. 791–800. [Online]. Available: http://doi.acm.org/10.1145/1526709.1526816

[4] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, H. Balakrishnan, and S. Madden, "Cartel: A distributed mobile sensor computing system," in *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems*, ser. SenSys '06. ACM, 2006, pp. 125–138. [Online]. Available: http://doi.acm.org/10.1145/1182807.1182821

[5] *3G Apple iOS Devices Are Storing Users' Location Data*, April 20, 2011.

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

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

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

[9] O. Hasan, B. Habegger, L. Brunie, N. Bennani, and E. Damiani, "A Discussion of Privacy Challenges in User Profiling with Big Data Techniques: The EEXCESS Use Case," in *Proceedings of the IEEE International Congress on Big Data*, June 2013, pp. 25–30.

[10] S. Schiaffino and A. Amandi, "Intelligent user profiling," in *Artificial Intelligence: An International Perspective*, M. Bramer, Ed. Berlin, Heidelberg: Springer, 2009, pp. 193–216. [Online]. Available: https://doi.org/10.1007/978-3-642-03226-4_11

[11] P. Samarati, "Protecting respondents' identities in microdata release," *IEEE Trans. Knowl. Data Eng*, vol. 13, no. 6, pp. 1010–1027, 2001. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/69.971193

[12] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 1, no. 1, 2007.

[13] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *Proceedings of the 29th IEEE Symposium on Security and Privacy*, 2008, pp. 111–125.

[14] ——, "De-anonymizing social networks," in *Proceedings of the 30th IEEE Symposium on Security and Privacy*. IEEE Computer Society, 2009, pp. 173–187. [Online]. Available: http://dl.acm.org/citation.cfm?id=1607723.1608132

[15] C. Dwork, F. Mcsherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *In Proceedings of the Third Theory of Cryptography Conference (TCC)*, ser. Lecture Notes in Computer Science, S. Halevi and T. Rabin, Eds., vol. 3876. Springer, 2006, pp. 265–284.

[16] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society, 2013, pp. 429–438. [Online]. Available: https://doi.org/10.1109/FOCS.2013.53

[17] Z. Qin, Y. Yang, T. Yu, I. Khalil, X. Xiao, and K. Ren, "Heavy hitter estimation over set-valued data with local differential privacy," in *Proceedings of the 23rd ACM Conference on Computer and Communications Security (CCS)*, E. R. Weippl, S. Katzenbeisser, C. Kruegel, A. C. Myers, and S. Halevi, Eds. ACM, 2016, pp. 192–203. [Online]. Available: http://dl.acm.org/citation.cfm?id=2976749

[18] Ú. Erlingsson, V. Pihur, and A. Korolova, "RAPPOR: randomized aggregatable privacy-preserving ordinal response," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, G. Ahn, M. Yung, and N. Li, Eds. ACM, 2014, pp. 1054–1067. [Online]. Available: http://doi.acm.org/10.1145/2660267.2660348

[19] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, and C. Palamidessi, "Broadening the scope of Differential Privacy using metrics," in *Proceedings of the 13th International Symposium on Privacy Enhancing Technologies (PETS 2013)*, ser. Lecture Notes in Computer Science, E. De Cristofaro and M. Wright, Eds., vol. 7981. Springer, 2013, pp. 82–102.

[20] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: differential privacy for location-based systems," in *Proceedings of the 20th ACM Conference on Computer and Communications Security (CCS 2013)*. ACM, 2013, pp. 901–914. [Online]. Available: http://doi.acm.org/10.1145/2508859.2516735

[21] K. Chatzikokolakis, E. ElSalamouny, and C. Palamidessi, "Efficient utility improvement for location privacy," *Proceedings on Privacy Enhancing Technologies (PoPETs)*, vol. 2017, no. 4, pp. 308–328, 2017.

[22] K. Fawaz and K. G. Shin, "Location privacy protection for smartphone users," in *Proceedings of the 21st ACM Conference on Computer and Communications Security (CCS 2014)*. ACM Press, 2014, pp. 239–250.

[23] K. Fawaz, H. Feng, and K. G. Shin, "Anatomization and protection of mobile apps' location privacy threats," in *Proceedings of the 24th USENIX Security Symposium, (USENIX Security 2015)*, J. Jung and T. Holz, Eds. USENIX Association, 2015, pp. 753–768. [Online]. Available: https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/fawaz

[24] L. Pournajaf, L. Xiong, V. Sunderam, and X. Xu, "Stac: Spatial task assignment for crowd sensing with cloaked participant locations," in *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '15. ACM, 2015, pp. 90:1–90:4. [Online]. Available: http://doi.acm.org/10.1145/2820783.2820788

[25] "Location guard," https://github.com/chatziko/location-guard.

[26] "QGIS Processing provider plugin (Methods for anonymizing data for public distribution)," https://github.com/SpatialVision/differential_privacy.

[27] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '11. New York, NY, USA: ACM, 2011, pp. 1082–1090. [Online]. Available: http://doi.acm.org/10.1145/2020408.2020579

[28] "The gowalla dataset." [Online]. Available: https://snap.stanford.edu/data/loc-gowalla.html

[29] P. Kairouz, S. Oh, and P. Viswanath, "Extremal mechanisms for local differential privacy," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 492–542, Jan. 2016. [Online]. Available: http://dl.acm.org/citation.cfm?id=2946645.2946662

[30] R. Agrawal, R. Srikant, and D. Thomas, "Privacy preserving olap," in *Proceedings of the 24th ACM SIGMOD International Conference on Management of Data (SIGMOD)*, ser. SIGMOD '05. New York, NY, USA: ACM, 2005, pp. 251–262. [Online]. Available: http://doi.acm.org/10.1145/1066157.1066187

[31] A. McIver, L. Meinicke, and C. Morgan, "Compositional closure for bayes risk in probabilistic noninterference," in *Proceedings of the 37th International Colloquium on Automata, Languages and Programming (ICALP 2010).*, ser. Lecture Notes in Computer Science, S. Abramsky, C. Gavoille, C. Kirchner, F. M. auf der Heide, and P. G. Spirakis, Eds., vol. 6199. Springer, 2010, pp. 223–235. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-14162-1

[32] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith, "Additive and multiplicative notions of leakage, and their capacities," in *IEEE 27th Computer Security Foundations Symposium, CSF 2014, Vienna, Austria, 19-22 July, 2014*. IEEE, 2014, pp. 308–322. [Online]. Available: http://dx.doi.org/10.1109/CSF.2014.29

[33] R. Shokri, G. Theodorakopoulos, and C. Troncoso, "Privacy games along location traces: A game-theoretic framework for optimizing location privacy," *ACM Transactions on Privacy and Security*, vol. 19, no. 4, pp. 11:1–11:31, 2017. [Online]. Available: http://doi.acm.org/10.1145/3009908