doi:10.11959/j.issn.1000-436x.2017241

差分隐私技术研究进展

高志强，王宇涛
（武警工程大学信息工程系，陕西 西安 710086）

摘 要：随着大数据共享时代的到来，数据隐私保护问题也随之突显。自 2006 年提出以来，差分隐私技术在支持隐私保护的数据挖掘与数据发布方面得到了广泛研究。近年来，Google、Apple 等公司陆续将差分隐私技术应用于最新产品中，差分隐私技术再次成为学术界和产业界的焦点。首先，对传统集中式模型下的差分隐私技术进行综述，介绍了面向数据挖掘与数据发布的差分隐私技术。然后，着重对最新的基于本地差分隐私模型下的数据收集与数据分析进行阐述，涉及众包模型下的随机响应、BloomFilter、统计推断等技术。最后，对差分隐私技术面临的主要问题和解决方案进行总结。

关键词：差分隐私；数据发布；数据挖掘；机器学习；众包；隐私保护
中图分类号：TP391 **文献标识码**：A

Survey on differential privacy and its progress

GAO Zhi-qiang, WANG Yu-tao
(Department of Information Engineering, Engineering University of PAP, Xi'an 710086, China)

Abstract: With the arrival of the era of big data sharing, data privacy protection issues will be highlighted. Since its introduction in 2006, differential privacy technology has been widely researched in data mining and data publishing. In recent years, Google, Apple and other companies have introduced differential privacy technology into the latest products, and differential privacy technology has become the focus of academia and industry again. Firstly, the traditional centralized model of differential privacy was summarized, from the perspective of analysis of data mining and data released in the differential privacy way. Then the latest local differential privacy regarding data collection and data analysis based on the local model was described, involving crowdsourcing with random response technology, BloomFilter, statistical inference techniques. Finally, the main problems and solutions of differential privacy technology were summarized.

Key words: differential privacy, data publishing, data mining, machine learning, crowdsourcing, privacy protection

1 引言

作为信息安全领域的重要部分，隐私保护问题一直是政府部门、企业、科研机构重点研究的课题。众所周知，政府的各职能部门每天存储、处理着大量公民的个人敏感信息；商业、互联网企业等也每时每刻在收集者用户的大量个人数据，如位置轨迹信息、消费习惯记录、注册登录信息、通话记录等；科研机构的研究人员却经常面临着对隐私数据的研究而又无法获得隐私数据的窘境，用户个人想直接控制或访问远在“云端”的私有数据却一直不能对数据的安全完全放心，而政府部门、企业也在为大量数据的集中存储、处理无比头痛，一旦用户数据泄露，无论对于个人还是社会都是一场巨大的灾难。因此，大数据环境下，数据的开放与治理中，数据的隐私保护问题不容忽视，不仅需要政府部门出台“重拳”措施治理，更亟待隐私保护技术的强有力技术支撑。

隐私保护问题的研究由来已久，Dalenius[1]提出了针对隐私控制（private disclosure control）的定义。

收稿日期：2017-09-18
基金项目：国家自然科学基金资助项目（No.61402529）
Foundation Item: The National Natural Science Foundation of China (No.61402529)

2002 年，*k*-anonymity 算法[2]的提出为接下来基于等价类分组的匿名隐私保护算法及其改进模型奠定了基础，*l*-diversity[3]、*t*-closeness[4], (*α*, *k*)-anonymity[5]等不断完善着针对不同攻击者背景知识的匿名保护理论。直到 2006 年，Microsoft 的 Dwork 提出可以抵抗攻击者任意背景知识的差分隐私技术[6~10]（DP, differential privacy）成为新的研究热点。差分隐私保护技术可以提供严格可证明的隐私保护，不仅丰富了隐私保护理论研究的内涵，目前更已被应用于实际产品中，如 Apple 的 iOS10 中的输入法及搜索功能和最新的机器学习 API——CoreML[11,12]、Google 的 Chrome 浏览器中[13,14]、Samsung 的智能手机[15]等。

目前，差分隐私技术的研究主要分为 2 种。1) 针对集中式数据模型（也称为基于可信第三方数据管理者模型（trusted curator））的传统支持差分隐私保护的数据挖掘与数据发布技术，其中，可以分类为交互式与非交互式、集中式与分布式、动态与静态数据等。2) 针对本地差分隐私（LDP, local differential privacy）模型的支持差分隐私保护的数据收集与数据统计分析与深度挖掘技术，其中，涉及随机响应技术、BloomFilter、统计分析、机器学习等技术。众包模式下的本地差分隐私保护技术之所以被产界和学界广泛认可，因为其不需要依赖于可信第三方数据管理者，用户数据的收集只涉及数据加噪音版本，原始真实数据完全被保护在本地设备，这既解决了用户对个人隐私数不能自主控制的关切，也降低了大量隐私数据在非可信第三方存储的隐私泄露风险。目前，LDP 技术已被应用于流式频繁项挖掘、基于众包的字符串统计估计、Google 的 Chrome 用户数据收集等领域。尤其，2016、2017 年 WWDC 大会上[11,12]，Apple 都将结合本地差分隐私的新技术应用于最新产品中，强调用户数据隐私的重要性，保证用户的隐私权益。

本文结合差分隐私技术的最新研究成果，首先，对传统集中式模型下的差分隐私技术进行综述，分析了面向数据挖掘与数据发布的差分隐私技术，重点对最新的基于本地差分隐私模型下的数据收集与数据分析进行阐述。最后，对差分隐私技术面临的主要问题和解决方案进行总结。

2 差分隐私保护模型

随着差分隐私保护技术研究的深入，在传统集中式可信管理者差分隐私保护模型（trusted curator）的基础上，针对众包模式的本地隐私保护模型（local model）成为差分隐私保护领域的有力补充。其中，本地模型直接在数据收集阶段将数据在用户本地进行隐私处理，从根源上保护数据隐私，避免了理想可信第三方介入，数据收集者在加噪版的数据基础上进行统计学习；而传统的集中式模型，数据收集者集中管理原始数据，统一对外提供支持差分隐私的数据发布、数据查询、数据挖掘接口。

由上述分析，可以得到集中式差分隐私与本地差分隐私的差异，其具体定义如下。

定义 1 (*ε*,*δ*)-*DP*[10]。若随机算法 F 满足(*ε*, *δ*)-*DP*，当且仅当所有在邻接数据库 D 和 D' 中，算法 F 的所有可能输出 $R \subseteq Range(F)$ 满足

$$\Pr(F(D) \in R) \leqslant \mathrm{e}^{\varepsilon} \Pr(F(D') \in R) + \delta \quad (1)$$

其中，ε 为来调节算法 F 输出隐私保护程度的参数，对于集中式差分隐私模型和本地差分隐私模型均适用。

定义 2 (*ε*,*δ*)-*LDP*[16]。若随机算法 F 满足(*ε*, *δ*)-*LDP*，当且仅当用户端数据对 v_1 和 v_2，对于算法 F 的所有可能输出 $R \subseteq Range(F)$ 满足不等式

$$Pr(A(v_1) \in R) \leqslant \mathrm{e}^{\varepsilon} Pr(F(v_2) \in R) + \delta \quad (2)$$

当 δ=0 时，定义 2 转化为 *ε*-*LDP*。通过调节隐私预算参数 ε，可以保证当户端数据改变时，数据收集者收到用户真实数据信息量改变不大。此外，在本地模型中 D 代表一个用户的数据，D'代表同一用户的依概率改变后的数据。而集中式模型的 D 代表所有用户的数据，D'代表除去有数据变化用户的所有用户数据。

针对定义 1 与定义 2 及上述分析，可以得到在差分隐私下的集中式与本地模型，如图 1 所示。

定义 3 序列组合特性[9]。存在 t 个随机算法 $A_i(1 \leqslant i \leqslant t)$满足 ε_i-*DP*，那么序列 $A_i(D)$满足$(\sum_{i=1}^{t} \varepsilon_i)$-*DP*。

差分隐私的序列组合特性是最常用的隐私预算 ε 分配策略（并行策略参考文献[15]）。

定义 4 敏感度[8]。有任意函数 f，敏感度 Δf 定义如下

$$\Delta f = \max_{D,D'} \left\| f(D) - f(D') \right\|_1 \quad (3)$$

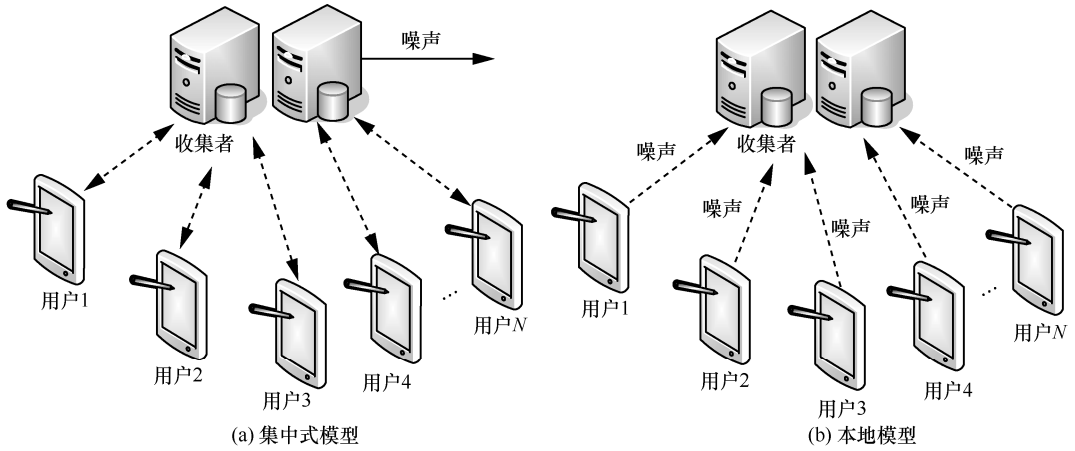

图 1　2 种隐私保护模型

其中，D 和 D'是邻接数据库，$\|\ \|_1$ 是向量的 ℓ_1 范式。在本地模型中，D 和 D'可以具体化为 2 条任意记录或 2 个二进制字符串。

定理 1　Laplace 机制[7]。函数 f：$D\to R^d$，敏感度为 Δf，随机算法 $A(D)=f(D)+Y$ 满足 ε-DP，其中，$Y\sim Lap(\Delta f/\varepsilon)$为随机噪声。

Laplace 机制是经常被用于本地模型中，常用于对数值型结果的隐私保护（指数机制、几何机制参考文献[9,10]）。

定义 5　随机应答[17]（RR, randomized response）是一种被用来保护敏感话题调查参与者隐私的技术。例如，每个人不是属于组 A 就是组 B，问题是在不能确定具体个人属于哪组的前提下，估计组 A 中人数的比例。

随机应答给出的解决方案：随机选取 n 个人，随机设备（可以是抛硬币、摸球模型）以概率 p 指向 A，以概率$(1-p)$指向 B。在每轮调查中，受访者只需回答设备指向（调查者未知）是否与其真正的组别一致（Yes 或 No），这样便可以得到组 A 人数的最大似然估计。通过证明分析，可以得到 A 真正比例 π 的无偏估计量[17]。此外，重要的是，RR 机制满足差分隐私机制，不依赖于攻击者的先验知识，可以在数据收集中保护任一参与者隐私，参与者可以拥有 $\varepsilon=\ln\dfrac{0.75}{1-0.75}=\ln 3$ 的隐私保护水平[13,14]。

3　差分隐私主要研究方向

3.1　面向数据挖掘与数据发布的差分隐私技术

随着互联网技术的兴起与大数据产业的发展，使面向数据挖掘与数据发布的差分隐私技术，近年来的研究在理论上不断发展和完善，并在统计学、机器学习、数据挖掘、社交网络等领域得到了初步应用。

从隐私控制的定义[1]到经典数据脱敏方法 k-anonymity[2]及其改进模型[3~5]，都无法克服 3 个方面缺点：基于可信第三方数据管理者；安全性严重依赖于攻击者所掌握的背景知识；无法提供严格且有效数学理论来证明其隐私保护水平。传统的集中式模型基于可信第三方，用户终端与数据收集者被视为一个整体，数据服务器存储未处理的原始用户隐私数据，经过隐私处理（如加噪）后统一对外发布。

目前，针对集中式差分隐私保护模型已有大量的研究成果[6~10,18~21]。Roth 等[18]提出了交互式数据发布的中位数机制（median），其能够在相同预算下提供更多数量的查询。Xu 等[19]提出了一种基于 k-d 树的直方图发布算法，当参数（频数分布紧密度阈值、空间分割次数）的取值适当时，DPCube 算法在查询数量和查询误差等方面具有更好的性能。Engel 等[20]提出的小波变换方法、Hay 等[21]提出的层次查询方法等。然而，这些针对差分隐私的数据发布和分析技术都基于可信管理者模型数据分布模型，集中式数据管理不可避免地面临着巨大的隐私安全风险。

3.2　基于本地差分隐私模型下的数据收集与数据分析

在 2013 年，Duchi[16]首先提出了 local differential privacy，而 Google 的 Chrome 浏览器的 RAPPOR[13]（randomized aggregatable privacy-preserving ordinal response）采用随机应答策略和 BloomFilter 实现了

针对客户端群体的类别、频率、直方图和字符串类型统计数据的隐私保护分析，可以提供 ln3 的差分隐私保护。在 RAPPOR 中，采用 2 个满足差分隐私的机制：永久和即时的随机响应，可以单独调节隐私保护水平，而且 BloomFilter 可以增加额外的不确定性，不仅压缩报文大小，更增加攻击者的攻击难度。在解码过程中结合成熟的假设检验、最小二乘求解和 LASSO 回归实现了针对字符串抽样群体频率的高可用解码框架。此外，RAPPOR 的改进模型实现了数据字典未知情况下的本地学习多变量联合概率分布估计。

针对流式频繁项挖掘问题，文献[22]在 RAPPOR 机制和 Succinct Histgram 的基础上，提出的 LDPMiner 方法，将挖掘任务分成 2 个子处理过程：Sampling SH 算法完成对流式频繁项的主成分识别工作，从噪声数据中初步确定流式频繁项的选值范围和 Sampling RAPPOR 算法对前一过程的结果进行频数估计上的调优处理，得到相比单一处理过程更为精确的流式频繁项结果。值得关注的是，Apple 是唯一一家将差分隐私作为标准大规模部署的公司，在 WWDC2016、2017 一直讲 LDP 技术应用于其最新产品中，主要涉及 3 个方面技术[11,12]：局部抽样、散列加密、噪声扰动。证明了统计学定理，只要数据量充足，即使只有是加噪音的数据，依然能建立起宏观视角统计大规模群体表现出来的倾向。

针对 LDP 技术的差分隐私理论分析主要涉及统计分析理论、差分隐私证明等。例如，可证明 RAPPOR 满足差分隐私的定义。其中，永久随机响应（PRR）保证了来自真值的加噪值保护隐私，可证明 RAPPOR 中 PRR 满足差分隐私[13]，同时，即时随机响应（IRR）满足差分隐私[14]。

4 结束语

本文在差分隐私的理论基础上，分别对集中式和本地模型下差分隐私技术在数据挖掘、数据发布、数据收集、数据分析、理论证明等方面进行综述。着重对 LDP 技术进行分析讨论，可以看到，LDP 技术近年来在互联网领域的大规模应用给学术界和产业界都带来强大动力。但对于差分隐私保护，在理论和应用上都还存在一些难点以及新的方向需要进一步深入研究，包括基于差分隐私的众包机器学习；抵抗新型攻击的能力，尤其是在生成式对抗网络[23]（GAN）环境中需要研究者引起注意；基于大数据平台 Hadoop、Spark、Storm 等的差分隐私下大数据分析等。

参考文献：

[1] DALENIUS T. Towards a methodology for statistical disclosure control[J]. Statistic Tidskrift, 1977, 15(2):429-444.

[2] SWEENEY L. *K*-anonymity: a model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570.

[3] MACHNAVAJJHALA A, GEHRKE J, KIFER D, et al. *l*-diversity: privacy beyond *k*-anonymity[C]// International Conference on Data Engineering. IEEE, 2006:24.

[4] LI N H, LI T C, VENKATASUBRAMANIAN S. *t*-closeness: privacy beyond *k*-anonymity and *l*-diversity[C]// IEEE International Conference on Data Engineering. IEEE, 2007:106-115.

[5] WONG R C, LI J Y, FU A W, et al. (*α*, *k*)-anonymity: an enhanced *k*-anonymity model for privacy preserving data publishing[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2006:754-759.

[6] DWORK C. A firm foundation for private data analysis[J]. Communications of the ACM, 2011, 54(1):86-95.

[7] DWORK C, KENTHAPADI K, MCSHERRY F, et al. Our data, ourselves: privacy via distributed noise generation[C]//International Conference on the Theory and Applications of Cryptographic Techniques. 2006:486-503.

[8] DWORK C, MCSHERRY F, NISSIM K. Calibrating noise to sensitivity in private data analysis[M]//Theory of Cryptography. Springer Berlin Heidelberg, 2006:637-648.

[9] DWORK C, NAOR M, PITASSI T, et al. Differential privacy under continual observation[J]. Stoc, 2010:715-724.

[10] DWORK C, NAOR M, PITASSI T, et al. Pan-private streaming algorithms[C]//The First Symposium on Innovations in Computer Science. 2010: 66-80.

[11] NOVAC O C, NOVAC M, CORDAN O, et al. Comparative study of Google, Android, Apple iOS and Microsoft Windows Phone mobile operating systems[C]//International Conference on Engineering of Modern Electric System. 2017: 154-159.

[12] SLVA M R D, ROMOS T M, HOLANDA M T D. Geographic information system with public participat on iOS system[C]//Information Systems and Technologies. 2017: 1-5.

[13] PIHUR V, KOROLOVA A. RAPPOR: randomized aggregatable privacy-preserving ordinal response[C]//ACM Sigsac Conference on Computer and Communications Security. ACM, 2014:1054-1067.

[14] FANTI G, PIHUR V, ULFAR E. Building a RAPPOR with the unknown: privacy-preserving learning of associations and data dictionaries[J]. Proceedings on Privacy Enhancing Technologies, 2016, 2016(3):41-61.

[15] NGUYEN T T, XIAO X, YANG Y, et al. Collecting and analyzing data from smart device users with local differential privacy[J]. arXiv: 1606.05053.

[16] DUCHI J C, JORDAN M I, WAINWRIGHT M J. Local privacy and statistical minimax rates[C]//Annual IEEE Symposium on Foundations of Computer Science. 2013: 429-438.

[17] WARNER S L. Randomized response: a survey technique for eliminating evasive answer bias[J]. Journal of the American Statistical Association, 1965, 60(309):63-66.

[18] DWORK C, ARON R. The algorithmic foundations of differential privacy[M]. Now Publishers Inc. 2014.

[19] XU J, ZHANG Z, XIAO X, et al. Differentially private histogram publication[C]//International Conference on Data Engineering. IEEE, 2012:32-43.

[20] ENGEL D, EIBL G. Wavelet-based multiresolution smart meter privacy[J]. IEEE Transactions on Smart Grid, 2017, 8(4):1710-1721.

[21] HAY M, MACHANAVAJJHALA A, MIKLAU G, et al. Principled evaluation of differentially private algorithms using DPBench[J]. arXiv: 1512.04817.

[22] QIN Z, YANG Y, YU T, et al. Heavy hitter estimation over set-valued data with local differential privacy[C]//ACM Sigsac Conference on Computer and Communications Security. ACM, 2016: 192-203.

[23] HITAJ B, ATENIESE G, PEREZ F. Deep models under the GAN: information leakage from collaborative deep learning[J]. arXiv: 1702.07464.

作者简介：

高志强（1989-），男，黑龙江齐齐哈尔人，武警工程大学博士生，主要研究方向为隐私计算、深度神经网络、群智能优化等。

王宇涛（1989-），男，贵州贵阳人，武警工程大学硕士生，主要研究方向为大数据挖掘。