

差分隐私的隐私观与隐私目标

贾靖宇^{1,2}, 谭畅^{1,2}, 刘哲伟^{2,3}, 李鑫豪^{2,3}, 刘哲理^{2,3}, 张涛⁴

(1. 南开大学计算机学院, 天津 300350; 2. 南开大学数据与智能系统安全教育部重点实验室, 天津 300350;
3. 南开大学网络空间安全学院, 天津 300350; 4. 中国工程物理研究院激光聚变研究中心, 四川 绵阳 621900)

摘要: 为了解决差分隐私中“隐私目标”难以理解的问题, 研究差分隐私在多个领域中存在的隐私争议。从数据相关性场景的具体示例展开, 研究学者对差分隐私保护目标的不同观点。当数据集中的记录具有相关性时, 敌手能利用记录间相关性从差分隐私机制中准确推理用户的某些敏感信息。这种现象是否违反隐私保护, 引起学者的长期讨论。借助法学领域的研究, 调研计算机领域两种主流隐私理论对隐私定义的影响。限制访问个人信息理论强调阻止他人获取个人的敏感信息。该理论认为隐私机制应当阻止敌手通过获取用户的真实信息。对个人信息的控制理论则强调个人向他人传达有关自身信息的权利。该理论认为因他人分享数据的相关性, 个人的信息泄露不应该被视为隐私泄露。分析计算机科学、社会科学、伦理道德以及人机交互领域中学者对于隐私概念的不同理解而对差分隐私产生的争议。从多学科视角分析差分隐私的隐私观, 帮助读者正确认识差分隐私的隐私观与隐私目标, 并增进读者对“隐私”的理解。

关键词: 差分隐私; 数据相关性; 隐私风险; 隐私理论

中图分类号: TP3-05

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2023071

Privacy view and target of differential privacy

JIA Jingyu^{1,2}, TAN Chang^{1,2}, LIU Zhewei^{2,3}, LI Xinhao^{2,3}, LIU Zheli^{2,3}, ZHANG Tao⁴

1. College of Computer Science, Nankai University, Tianjin 300350, China

2. Key Laboratory of Data and Intelligent System Security, Ministry of Education, Nankai University, Tianjin 300350, China

3. College of Cyber Science, Nankai University, Tianjin 300350, China

4. Laser Fusion Research Center, China Academy of Engineering Physics, Mianyang 621900, China

Abstract: The study aimed to address the challenges in understanding the privacy goals of differential privacy by analyzing the privacy controversies surrounding it in various fields. It began with the example of data correlation and highlighted the differing perspectives among scholars regarding the targets of privacy protection. In cases where records in a dataset were correlated, adversaries can exploit this correlation to infer sensitive information about individuals, thereby sparking a debate on whether this violates privacy protection. To investigate the influence of privacy theories in the legal domain on defining privacy, two mainstream privacy theories in the computer field were

收稿日期: 2022-11-30; 修回日期: 2023-03-31

通信作者: 刘哲理, liuzeli@nankai.edu.cn

基金项目: 国家自然科学基金(62032012)

Foundation Item: The National Natural Science Foundation of China (62032012)

引用格式: 贾靖宇, 谭畅, 刘哲伟, 等. 差分隐私的隐私观与隐私目标[J]. 网络与信息安全学报, 2023, 9(5): 82-91.

Citation Format: JIA J Y, TAN C, LIU Z W, et al. Privacy view and target of differential privacy[J]. Chinese Journal of Network and Information Security, 2023, 9(5): 82-91.

examined. The first theory, limited access to personal information, focuses on preventing others from accessing an individual's sensitive information. According to this theory, privacy mechanisms should aim to prevent adversaries from accessing a user's actual information. In contrast, the second theory, control over personal information, emphasizes an individual's right to communicate personal information to others. This theory suggests that the disclosure of personal information due to the relevance of others sharing data should not be considered a breach of privacy. Then the controversies of differential privacy were analyzed in the fields of computer science, social science, ethics and human-computer interaction due to their different understandings of privacy. By exploring the privacy concept of differential privacy from a multidisciplinary perspective, this study helps readers gain a correct understanding of the privacy viewpoint and goals of differential privacy while enhancing their understanding of the concept of "privacy" itself.

Keywords: differential privacy, data correlation, privacy risk, privacy theory

0 引言

数据已经成为当今社会上技术创新背后的关键驱动力，企业能够依靠海量数据提供更有竞争力的产品。因此，企业一直在促进各方共享数据以实现更大的商业以及社会价值。然而，数据隐私泄露事件的披露导致人们开始担心分析敏感数据的研究工作将对个人隐私造成的不利影响^[1-3]。人们迫切需要在数据分享的过程中提供足够的隐私保护，以规避不必要的隐私风险。为了促进数据共享，研究人员提出了差分隐私^[4-10]（DP，differential privacy）的隐私保护概念。

差分隐私是数据分析场景下隐私保护的黄金标准，在指导隐私算法的设计以及量化数据分析用户的隐私泄露水平任务中发挥着重要的作用。差分隐私要求隐私算法应当确保单个记录的变化几乎不会影响数据分析的结果。在这种情况下，敌手不太可能从差分隐私保护下的分析结果中识别单个用户的记录。因此，无论是用户还是企业在共享数据时，都不必过分担心数据共享行为对个人隐私造成太大的风险。

随着差分隐私的迅速发展，一系列关于差分隐私保护目标的争论^[11-19]开始出现。学者发现很难解释差分隐私在现实场景上提供的隐私保护效果。尽管差分隐私拥有严谨的数学属性，能够在复杂的场景中量化隐私泄露，但差分隐私的量化指标并不直观，需要结合上下文来理解，这使得学者在差分隐私的保护目标与适用场景上产生了分歧。

在数据存在相关性的场景中，差分隐私的争议引发了学术界对于隐私目标的大量讨论。2011

年，Kifer 等^[12]指出如果原始数据集的记录间存在相关性，差分隐私很难提供有实际意义的保护效果。一个拥有辅助信息的敌手能准确推理差分隐私保护的用户敏感信息。只有假设记录相互独立，差分隐私提供的保护才是有意义的。基于 Kifer 等的研究，相当数量的学者认为如果数据间存在相关性，使用差分隐私算法会导致用户的隐私泄露^[20-28]。对于这一问题，争论的关键并不在于技术，而是如何正确理解“隐私”这一概念。

在统计数据分析中，传统的隐私目标是在统计数据中学习用户特征的同时，防止敌手推理出某些个体的属性。然而，不断出现的隐私泄露事件表明，敌手总是能借助特定的辅助信息来有效地从统计信息中推理出一些个人属性。在任意背景知识的敌手下同时兼顾统计数据的效用和隐私是不现实的。相对应地，差分隐私假定无法阻止个人信息的泄露。但为了促进用户分享数据，差分隐私要求隐私保护技术在数据分析结果中提供一定的随机性以保证单个记录的变化不会对分析结果产生较大的影响。换句话讲，差分隐私保证即使出现隐私泄露的情况也与用户分享数据的行为无关。然而，差分隐私的隐私目标并不符合所有人对隐私的理解。学术界关于隐私目标的争议与混淆开始阻碍差分隐私技术的发展。

本文认为，提高用户对“隐私”概念的理解是一项很重要但往往被忽略的工作。在科研工作中，常常能看到“侵犯隐私”或者“隐私泄露”的描述，但学者往往没有对他们声称的“隐私”进行明确的定义。而学者对于隐私这一概念的理解，很可能会影响他们文章的立意。当隐私保护的目标与人们的

隐私观念冲突时,非技术的不可调和的矛盾就会出现。在提到隐私保护时,研究人员通常关注隐私算法的技术特征,而忽略这些隐私算法在“隐私”理解上可能存在的差异。而在现实场景中,作者的隐私观对隐私技术的落地有重要意义。

为了解释差分隐私在学术界引发的一系列争议,并帮助读者理解差分隐私的隐私观和隐私目标,本文总结关于差分隐私的一系列争议,并解释差分隐私场景下的隐私目标。同时,本文引入法学界对于“隐私”这一概念的长期讨论,增进读者对“隐私”概念的理解,并帮助读者正确理解差分隐私的隐私目标,从而正确地学习和使用差分隐私以及其他隐私保护技术。

1 差分隐私

差分隐私是一种具有规范数学性质的隐私定义,它要求隐私算法在数据分析结果中提供足够的随机性以保护用户隐私。差分隐私承诺如果用户得到差分隐私算法的保护,用户所面临的隐私泄露风险与他们分享数据的行为无关。差分隐私认为,当敌手拥有足够的背景知识时,用户面临的信息泄露风险是不可避免的。但差分隐私通过为用户提供适当的保护以降低数据分享与敏感信息泄露的关联性。

1.1 差分隐私定义及相关概念

定义 1 差分隐私^[5]。设 $\varepsilon \geq 0, \delta \in [0,1]$, 如果输出域为 $\mathcal{Y}$ 的隐私算法 M 满足 (ε, δ) -差分隐私, 那么对于任意相邻数据集 (只有一条记录不同的两个数据集) $X, X' \in \mathcal{X}^n$, 以及任意的输出域 $Y \subseteq \mathcal{Y}$, 均有

$$\Pr[M(X) \in Y] \leq e^\varepsilon \Pr[M(X') \in Y] + \delta$$

其中, 参数 e 为自然常数。参数 ε 为隐私预算, 指用户的数据在经过差分隐私算法保护后的隐私损失上界, 该参数越大, 则隐私算法的泄露风险越高。参数 δ 表示隐私损失超过参数 ε 约束的概率 (通常被设置为一个可忽略的数值)。隐私算法 M 在相邻数据集 X, X' 上的输出差异将被隐私参数 (ε, δ) 约束, 当参数 (ε, δ) 较小时, $M(X)$ 和 $M(X')$ 的分布几乎相同, 即单个记录的改变不会影响数据分析的结果。差分隐私不考虑敌手从算法输出得到的计算优势, 而是度量算法本身的随

机性。因此, 只要差分隐私算法能够正确执行, 用户就能得到对应的隐私保护, 敌手的辅助信息不会破坏隐私保护的效果。

接下来, 本文给出全局灵敏度的定义。全局灵敏度指的是单条记录对查询函数的影响程度, 它用于指导不同查询函数中差分隐私机制的噪声尺度。

定义 2 全局灵敏度^[5]。给定函数 $f: \mathcal{X}^n \rightarrow \mathbf{R}^d$, 对于任意相邻数据集 $X, X' \in \mathcal{X}^n$, 查询函数 f 的全局灵敏度为

$$\Delta f = \max_{X, X'} \|f(X) - f(X')\|_p$$

其中, $\|\cdot\|_p$ 表示 l_p 范数。

下面介绍一种常用的差分隐私机制——拉普拉斯机制。拉普拉斯机制通过在查询结果中添加服从拉普拉斯分布的随机噪声提供差分隐私保护。

定义 3 拉普拉斯机制^[5]。对于任意查询函数 $f: \mathcal{X}^n \rightarrow \mathbf{R}^d$, 以及任意数据集 $X \in \mathcal{X}^n$, 拉普拉斯机制 $M(X) = f(X) + (\eta_1, \dots, \eta_d)$ 满足 $(\varepsilon, 0)$ -差分隐私, 其中 η_i 为服从拉普拉斯分布 $\text{Lap}(\Delta f / \varepsilon)$ 的随机变量。

以满足 $(1, 0)$ -差分隐私的拉普拉斯机制为例: 假设有一个计数查询任务, 要求计算某药店某日的购药人数。为了保护用户的购药信息, 药店使用拉普拉斯机制在查询结果中添加噪声。假设当日的购药人数为 10, 那么拉普拉斯机制在相邻数据集上的输出分布如图 1 中蓝线所示; 而如果当日的购药人数为 11, 机制在相邻数据集上的输出分布则如图 1 中橙线所示。由于两个分布在任意结果上都有着相近的输出概率, 因此敌手很难准确推断当日的购药人数, 从而保护用户隐私。

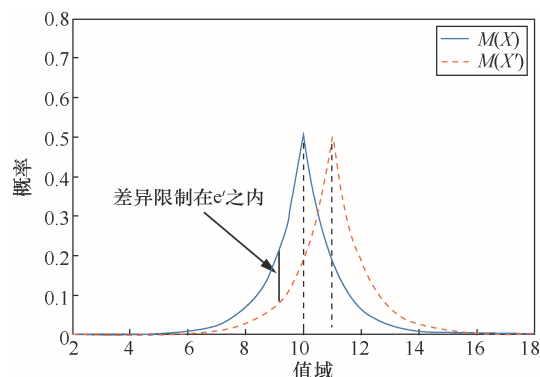


图 1 拉普拉斯机制在相邻数据集上的输出分布
Figure 1 Distribution of the Laplace mechanism on neighboring datasets

1.2 统计披露限制与差分隐私

数据分析的隐私保护问题跨越了多个学科。越来越详细的个人电子数据和不断发展的数据分析技术对数据的收集和管理提出了更高的要求。在差分隐私出现之前, 隐私保护技术往往会对一个直观的隐私指标进行保护。以 k -匿名化^[29]为例, 该技术要求发布的表格式数据集中相同的组合需要出现至少 k 次。然而, 如果原始数据集中记录缺乏多样性, 敌手很容易通过统计结果推理出某个人的属性。由于缺少统一的设计标准, 隐私算法很容易在不同背景知识的攻击者下遭到攻击。因此, 学术界需要一个能够抵御任意背景知识敌手、有实际应用意义, 并且有着严格数据性质的隐私定义来指导隐私保护技术的设计并量化隐私技术的保护效果。

1977年, Dalenius^[30]对统计数据库提出了个人隐私保护要求: 除非能直接接触数据库, 关于个人的任何信息都不应该从数据库的统计信息中学习。Dalenius 的隐私目标也被称为经典的统计披露限制。统计披露限制认为, 如果敌手能从统计信息中学习到关于个人的部分信息, 统计数据库就是不隐私的。类似的安全性定义常见于现代密码学, 但很难匹配一些数据分析的实际需求。

在2008年的一项研究中, Dwork^[31]用一个简单的例子指出统计披露限制的困境: 假设数据库能计算不同地区女性的平均身高。只要敌手有特定的辅助信息, 统计披露限制的隐私目标就几乎不可能实现, 除非统计结果中几乎不包含任何有用的信息。

1.3 差分隐私的隐私观

由于敌手辅助信息存在, 合理限制敌手的推理能力并提供有意义的隐私保护是较为困难的。同时, 如果希望限制敌手辅助信息的类型, 如何界定敌手的辅助信息是否合理就会成为一个困难的问题。为了提出一个合理的隐私定义, 差分隐私选择对“隐私”进行狭义解释, 通过降低隐私保护的标准而避免出现上述问题^[5]。

差分隐私中, 用户只应当对他们能控制的行为要求隐私保护。因此, 差分隐私的目标是合理量化用户分享数据的这一行为可能导致的隐私泄

露。差分隐私限制分享数据导致的信息泄露, 而不限实际敌手推理导致的数据泄露程度, 因此它并不需要对敌手的辅助信息进行假设。作为一个对隐私机制随机性的量化标准, 差分隐私自然地抵御任意背景知识敌手。即使敌手拥有任意的背景知识, 能以极高的概率从数据分析中推理用户的真实信息, 也并不影响差分隐私的保证。

2 差分隐私的争议

差分隐私对“隐私”概念的狭义解释使得它面对任意背景知识敌手都能稳定地量化用户的隐私泄露。由于其规范的数学属性, 以及能抵御任意敌手的强鲁棒性, 差分隐私在敏感信息收集、数据库隐私保护、医疗数据分析、机器学习隐私保护等场景得到了广泛应用。然而, 一些研究表明差分隐私算法可能无法提供足够的隐私^[12,20-28]。Kifer^[20]认为, 差分隐私模糊了人们隐私需求中试图保护的内容。具体来讲, 差分隐私只保护个人记录在数据集中的变化所造成的影响, 但实际上, 它并不能完全防止敌手从数据分析任务能推理出个人信息, 而这一点往往与人们对隐私的直观理解不符。本文借助数据具有相关性的场景中差分隐私的争议来解释差分隐私可能存在的隐私问题。

2.1 差分隐私与数据相关性

本文通过传染病信息发布案例来解释差分隐私的保护目标以及学者争议的焦点。差分隐私数据发布示意如图2所示, 展示了用户敏感信息在数据收集与发布过程的生命周期。假设有一种传染病会感染整个家庭, 共有 n 个用户参与了传染病信息的收集, 其中包括Bob家族的9个人。如图2所示, 数据管理员收集到所有用户上传的信息后, 通过满足 $(1,0)$ -差分隐私的拉普拉斯机制发布患病人数的统计信息。图中的每个箭头表示箭头两端之间的关系, 这些关系可以是一些已知的信息、某种未知的相关性或数据流的方向。箭头1表示家庭成员的感染状态(患病或未患病)。箭头2代表Bob的家庭成员和社会之间的一组关联。箭头3、4和5表示状态集中的每个状态对应数据集中的原始数据点。箭头6、7和8表示以数据集 D 作为输入的拉普拉斯机制。

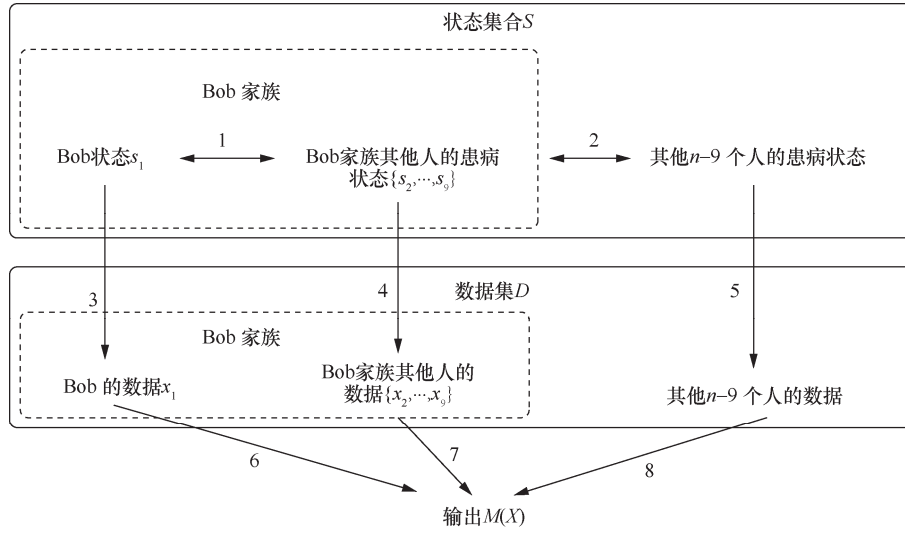


图2 差分隐私数据发布示意
Figure 2 Illustration of differential privacy data release

假设对手的目标是推断用户 Bob 的敏感信息。敌手可以结合其背景知识和统计信息来推断用户的状态。

在图2的数据收集示例中, Bob 的患病状态信息是如何流动的? 首先, 路径3-6表示 Bob 自身参与数据收集导致其患病状态的泄露; 路径1-4-7表示 Bob 家族其他人参与数据收集对 Bob 患病信息泄露的影响。如果 Bob 家族的患病情况与其他用户还有额外的相关性, 路径1-2-5-8同样会在一定程度上泄露 Bob 的患病状态。

假设敌手的背景知识包括除了 Bob 家族其他人以外的全部信息、传染病会在家庭中传播的特征, 以及 Bob 家族参与数据收集的人数, 考虑实验1所示的推理攻击实验: 将敌手的推理算法 A 、隐私算法 M 以及潜在的两个数据集 X, X' 作为输入, 推理隐私算法的输出 y 来自哪个数据集。

实验1 推理攻击实验

- 1) $\text{Exp}(A, M, X, X')$
- 2) 输入: 敌手的推理算法 A , 差分隐私算法 M , 数据集 X, X'
- 3) 均匀随机从 $\{0, 1\}$ 中选择变量 b
- 4) 当 $b=0$ 时, 构造 $S=X$
当 $b=1$ 时, 构造 $S=X'$
- 5) 差分隐私算法 M 计算 $y=M(S)$
- 6) 如果攻击者在获取 y 后推理 $A(y)$
- 7) 输出: 如果 $A(y)=b$, 即敌手成功推理输

出来自哪个数据集返回 1; 否则返回 0。

由于背景知识的存在, 敌手清楚潜在的两个数据集对应着 Bob 家族其他人全部患病或全部健康。定义敌手推测成功的期望为 $E(\text{Exp}(A, M, X, X'))$, 后文简写为 $E(\text{Exp})$, 将隐私算法 M 的输出域 $\mathcal{Y}$ 分为不相交的域 y_0, y_1 , 使得 $y_0 \cap y_1 = \mathcal{Y}$ 。如果隐私算法的输出 $y \in y_0$, 则推测数据集 $S=X$, 否则推测 $S=X'$ 。敌手推测成功的期望如下所示。

$$\begin{aligned} E(\text{Exp}) &= P[A(y)=0, S=X] + P[A(y)=1, S=X'] = \\ &= P_0 \cdot \Pr[A(y)=0 | S=X] + P_1 \cdot \Pr[A(y)=1 | S=X'] = \\ &= P_0 \cdot \Pr[M(X) \in y_0] + P_1 \cdot \Pr[M(X') \in y_1] = \\ &= \frac{1}{2} (\Pr[M(X) \in y_0] + \Pr[M(X') \in y_1]) \end{aligned}$$

给定输出 y , 当 $\Pr[M(X)=y] > \Pr[M(X')=y]$ 时, 敌手推测数据集为 X , 否则数据集为 X' 。借助上述最大化推理成功概率的方法, 通过计算拉普拉斯机制的累积分布函数, 可以得知敌手有接近 99.5% 的概率正确推理 Bob 的患病信息。然而, 满足 (1,0)-差分隐私的拉普拉斯机制中, 相邻数据集的输出分布的统计差异不到 40.0%, 这意味着, 差分隐私的保证与敌手推理能力不匹配。

上述例子中, 敌手关于数据相关性的辅助信息能帮助其以更高的概率推理出 Bob 的患病状态。在图2所示的信息流中, 差分隐私只量化了路径3-6的信息流对 Bob 患病状态的泄露。然

而，如果用户希望阻止敌手推理用户的真实信息，那么不仅需要保护路径 3-6 的信息流，还需要保护路径 1-4-7 甚至是路径 1-2-5-8 信息流。

2.2 长期争议

针对上述问题，一些学者认为差分隐私在数据存在相关性的场景中并不能对个人的隐私泄露提供充分的保护^[12,20-28]。其中，一些工作认为差分隐私是建立在数据独立的假设之上的^[12,24,25]。只有数据集中所有记录相互独立时，差分隐私的参数才能有效地表示为敌手的推理能力。差分隐私认为敌手的推理能力并不能代表用户隐私泄露的程度。即使是在图 2 的例子中，敌手能以极高的概率推理 Bob 的患病信息，也不代表差分隐私没能保护用户的隐私。而他家人分享数据导致的信息泄露，不应当作为 Bob 的隐私来保护。在后续的研究中，Michael 等^[21]基于上述工作，提出通过关联和因果关系来解释上面两种相互矛盾的观点。Michael 指出，差分隐私应从因果关系的角度理解：差分隐私能保证敌手的推理能力与用户分享数据缺乏因果关系，但并不能保护数据集本身的关联性。

2019 年，哈佛大学组织召开了一场关于差分隐私的研讨会。差分隐私学者、计算机学者、经济学家与社会学家共同讨论并试图找出如何以及何时应用差分隐私的共识。在如何衡量隐私泄露上，学者始终没能达成统一的意见。差分隐私的研究人员希望将差分隐私作为隐私泄露的标准，但在场的研究人员则更关注如何评估差分隐私的保护效果，即在差分隐私的保护下，如何衡量敌手重识别用户数据的能力。由此看来，差分隐私在现实场景的应用还有较长一段距离。

本文认为有关差分隐私的争议本质上是学者对隐私有不同理解，进而导致对隐私保护产生了不同的期望。由于双方的矛盾是主观的，因此任何技术上的研究都无法解决学者在差分隐私上的争议。下文将描述学者的两类隐私观，并借助法学与社会学家的研究总结两类隐私观的特征。

3 信息隐私的主流隐私观

本文借助法学对隐私的长期研究，将两种不同的隐私理论呈现给读者。Himma 等^[32]基于对数

据信息和计算机伦理的研究，将与计算机数据有关的隐私定义为信息隐私，并指出“有关信息隐私的问题的大多数分析援引了限制访问理论和控制理论的变体。”本文发现，学者对差分隐私的不同理解，是这两种理论的延伸。

3.1 限制访问个人信息理论

限制访问个人信息（limited access to personal information）理论是最重要的隐私理论之一，它强调隐私是阻止他人获取个人信息的一种权利。Gavison^[33]认为“隐私是对他人访问个人信息的一种限制。”类似地，Bok^[34]认为“隐私是一种被保护的财产，不接受不希望的访问，包括物理访问、个人信息或注意。”在限制访问理论中，当一个人完全独立于社会而存在时，他就获得了绝对的隐私。显然，人们很难在社会中获得绝对的隐私或完全的隐私丧失。因此，限制访问个人信息理论关注的是“隐私的丧失”。Gavison^[33]认为“当一个人获得他人的信息、关注他们或获得访问他们的机会时，就会发生隐私的丧失。”Tavani 等^[35]也有类似的观点，认为“隐私从根本上是关于保护他人免受入侵和信息收集”。

限制访问理论是对隐私的一种直观感受：如果一个人不希望其他人访问他的信息，但其他人通过某种方式了解了他的信息，那么他的隐私就会受到侵犯。然而，实现在信息隐私中限制访问理论存在一定困难。Tavani^[35]指出：限制访问理论的深刻见解之一是认识到需要建立边界和上下文的重要性，以实现信息隐私。换句话说，如果希望在限制访问理论下保障隐私，就需要确定隐私的范围，以及与人们希望保障的隐私相关的各种类型的信息。然而，这正是限制访问理论面临的主要困境。限制访问理论缺乏对隐私信息和公共信息的明确区分，因此人们很难建立一个统一的标准来辨别何种程度的接触会构成侵犯隐私。

大部分密码学技术包括统计披露限制，均符合限制访问理论对于隐私保护的理解。这些工作通常将防止敌手获取用户的敏感信息作为隐私保护的最终目标。然而，由于难以界定哪些是用户的隐私信息，基于限制访问理论提供隐私保护尚存在诸多困难。例如，人物自传中通常会涉及与

他人的一些生活经历,然而在未恶意歪曲事实的前提下,作者可以在写作中合法透露他人的部分经历而不承担任何后果。随着科技的进步,在医疗方面面临着更加复杂的隐私问题。以基因隐私为例,家庭成员间的基因通常有着密切的关联性,而个人对基因信息的处置权利是否应当考虑家庭其他成员的隐私尚且处于法律和道德的空白领域。

3.2 个人信息控制理论

对个人信息的控制(control over personal information)理论是另一个重要的隐私理论。这一理论的提出者是美国隐私理论的创始人 Alan Westin。Westin^[36]认为“隐私是个人、团体或机构自行决定何时、如何以及在何种程度上向他人传达有关他们的信息的权利。”Westin 观察到,个人同时具有披露信息和他人陪伴的需求,这与他们的隐私需求一样重要。这一观察表明,人类需要在信息披露和个人隐私之间取得平衡。这个概念在现实生活中很常见。一个人有时可能想和家人或朋友待在一起,有时又想独处。然而,过多或过少的隐私会破坏平衡,严重影响人们的生活。因为每个人都需要不断地在独处和群集之间调整,Westin 认为:个人为自己决定何时应该向公众披露自己的信息是个人隐私权利的核心。

控制理论同样有着众多的支持者^[37-39]。Fried^[37]指出了限制访问理论和控制理论的区别,他认为“人们对我们的了解越少,我们的隐私就越多,这种说法是不对的。隐私不仅仅是在别人的脑海中没有关于我们的信息,而是我们对自身信息的控制。”

由于对隐私的狭义解释,控制理论受到了许多学者的批评。Tavani 等^[35]认为个人对敏感信息的控制是非常有限的,控制理论排除了很多应该被理解为隐私的东西。他们建议,即使人们失去了对敏感信息的控制,这些信息也应该作为隐私信息加以保护,如传染病发布问题以及基因隐私问题。个人的信息往往与他人具有关联性,而他人对信息的处置有可能导致自身的信息泄露。为了保护某些敏感信息,人们需要扩大隐私的范围,以保护某些不受自己控制的信息分享行为。

3.3 两种理论的矛盾

基于限制访问理论的统计披露限制和基于控制理论的差分隐私是目前数据分析场景中学者对于隐私保护的主流理解。这里总结两种隐私理解的差异。

首先,对于“隐私”的理解,统计披露限制认为隐私保护是为了阻止敌手推理用户的敏感信息;而差分隐私则认为隐私保护是为了量化用户分享数据行为对隐私泄露的影响。其次,对于用户面临的数据重识别风险,统计披露限制认为用户的隐私风险是可以被限制的,因此应当通过隐私技术来抵御此类风险,并通过适当的指标对风险进行评估,而差分隐私则认为隐私泄露的风险是不可预见且难以阻止的,因此不考虑敌手的攻击效果而只量化隐私算法的随机性,进而促进用户分享数据。

在量化隐私保护效果时,统计披露限制需要对敌手的推理能力进行限制。对于具有辅助信息的敌手,统计披露限制很难给出统一的量化标准,当敌手的辅助信息超出限制时将无法提供有意义的保护。差分隐私不考虑敌手的攻击效果,而是量化隐私算法的随机性,因此只要隐私算法能正确运行,差分隐私的保护在任意辅助信息的敌手下都是有效的。如果研究人员希望从统计披露限制的角度使用差分隐私技术,就需要设置敌手的能力以及直观的隐私泄露指标(如假设检验中的误报率与漏报率)来重新量化差分隐私技术的保护效果。

作为一种基于控制理论的隐私定义,差分隐私自然地继承了控制理论的特点。其对隐私的狭义解释招致了诸多学者的不满,并且在数据具有相关性的场景中,很难提供具有实际意义的保护。然而,差分隐私的广泛应用表示从信息控制的角度量化隐私泄露是今后隐私保护的大趋势。

正如 Nissim^[11]所述,人们对隐私保护的期望往往依赖于信息交互的直觉假设而不是基于科学与数学原理。以这种方式构建隐私可能出现不现实的隐私需求,导致学界追求一个不可能实现的理想化的隐私目标。相对地,纯技术方法可能会对隐私采取狭隘的观点,无法捕捉隐私的基本规范期望。如何合理地定义隐私,兼顾科学计算并

满足社会对隐私的一般期望，将会是一个复杂的研究方向。

4 其他领域中差分隐私的争议

近年来，针对差分隐私的争议不局限于计算机领域的技术研究。在与现行法律的结合、社会科学中人们对隐私的普遍期望以及人机交互研究中普通用户对差分隐私的理解等方面引起了学者的一定关注。

4.1 法律法规和社会科学

在《美国科学院院报》发表的一项工作中，Cohen^[13]对欧洲《通用数据保护条例》(GDPR)的隐私目标进行建模，将GDPR对隐私泄露的描述解释为：数据发布程序应当防止敌手以较高的概率从统计结果中找出一条记录与数据集中的真实记录匹配。他们将建模的隐私目标定义为防谓词单选安全(PSO secure, secure against predicate singling out)，并证明了差分隐私满足PSO安全。然而，Holzel^[17]则指出差分隐私并不能保护数据主体免受由于统计数据库的存在而造成的伤害，因为它不能保证敌手利用背景知识从发布的数据集中推理用户的敏感信息，从而违反GDPR的隐私目标。

在美国联邦贸易委员(FTC, Federal Trade Commission)对消费者隐私保护的研究报告^[40]中，FTC对企业保护个人数据隐私提出的要求中的第一条就是“企业必须采取合理措施，确保数据被去识别。这意味着企业必须达到合理的信心水平，不能合理地使用数据来推断特定消费者、计算机或其他设备的信息，或以其他方式与之关联。”如果企业使用差分隐私机制为个人数据提供隐私保护，FTC的要求意味着不能单纯地通过差分隐私的参数判断是否提供足够的保护，而需要考虑实际场景下可能敌手的推理能力。

针对学术界关于差分隐私的争议，Oberski^[14]从社会科学的角度审视了差分隐私。他认为差分隐私可能会彻底改变社会对传统隐私泄露的理解。Oberski认为关于差分隐私的技术性研究并不能解决人们在隐私目标上的分歧，因为分歧与科学事实无关。相反，人们对隐私风险有不同的主观信念，因此他们对如何减轻此类风险的想法不

同。如果不能统一人们对隐私的理解，关于差分隐私的分歧可能永远都无法解决。

4.2 人机交互领域

近几年开始出现针对差分隐私的人机交互研究，学者逐渐尝试向普通用户解释和推广差分隐私^[16,18-19]，这意味着差分隐私已经从纯技术性的研究发展到调研用户体验的新阶段。Xiong等^[18]在收集健康信息的应用程序中向外行人描述差分隐私提供的技术与保护。Xiong等发现提供差分隐私的描述能帮助用户提高分享敏感数据的意愿，然而调查中对客观理解问题的回答表明，普通用户很难理解差分隐私中对隐私的承诺和其他隐私保护技术的区别。在后续的工作中，Xiong等^[19]尝试通过各种描述方法改善人们对差分隐私的理解，在对400位具有大学教育背景和计算机专业知识的对象进行调查后，Xiong等发现调查对象对差分隐私的理解依旧不如预期。

Cummings等^[16]调研了普通用户对差分隐私的期望。他们发现当下企业对差分隐私不恰当的描述很有可能误导用户设置过高的隐私期望。Cummings等认为，对用户解释差分隐私不应仅仅为了提高他们分享数据的意愿，而应该帮助用户正确理解差分隐私能提供的保护。调查结果显示，企业对差分隐私的描述会改变用户对于隐私风险的评估，然而超过半数受访者关心的隐私泄露与差分隐私承诺的保护目标是不同的。

4.3 差分隐私的伦理研究

美国已经在人口普查中应用差分隐私在数据发布中保护公民隐私^[41]。谷歌^[6,7]、苹果^[8]、微软^[9]等企业也在努力拓展差分隐私的应用场景。在技术角度的研究中，学者通常会详尽列出所有可能出现的情况，并分析在各种情况下隐私保护技术的效果。而在人机交互的研究中，学者更加关注如何更好地帮助用户理解和使用即将(或已经)在实际场景中广泛部署的技术。因此，针对差分隐私的伦理研究还处于空窗期。尽管已经有学者意识到，广泛应用差分隐私技术可能会引发一系列的伦理道德甚至法律问题。但无论是计算机、人机交互还是法学研究领域都没能为差分隐私提供一个准确的适用范围。

5 结束语

本文总结了差分隐私在隐私观和隐私目标上的长期讨论,并从计算机科学、法学、人机交互等多个领域调研学术界在差分隐私上的争议。经过长期的研究,差分隐私逐渐成为一种成熟可靠的数学工具,在越来越多的领域中为用户提供隐私保护。然而,差分隐私对隐私的理解可能与人们对隐私的一般理解不同,这使得在实际部署和增进用户理解差分隐私时困难重重。差分隐私已经逐渐从一项技术性的研究转变为社会、法律、伦理道德与人机交互等多学科交叉的研究领域。因此,除了关于差分隐私技术性的科学研究之外,加深人们对“隐私”概念的深入理解是促进差分隐私应用落地的重要推手。

差分隐私已经在事实意义上成为收集、分析与发布用户数据的重要工具。在此情况下,确定数据分析场景的隐私保护的范,以及差分隐私的适用场景变成了一个关系到公民隐私保护的重要研究领域。设计一套行之有效的沟通机制,帮助普通用户理解差分隐私,并指导用户做出正确的数据分享决策是目前差分隐私领域有待深入挖掘的研究方向。

参考文献:

- [1] PRICE W N, COHEN I G. Privacy in the age of medical big data[J]. *Nature Medicine*, 2019, 25(1): 37-43.
- [2] WANG Z, SONG M, ZHANG Z, et al. Beyond inferring class representatives: User-level privacy leakage from federated learning[C]//IEEE INFOCOM 2019-IEEE Conference on Computer Communications. 2019: 2512-2520.
- [3] AYABURI E W, TREKU D N. Effect of penitence on social media trust and privacy concerns: The case of Facebook[J]. *International Journal of Information Management*, 2020, 50: 171-181.
- [4] DWORK C, MCSHERRY F, NISSIM K, et al. Calibrating noise to sensitivity in private data analysis[C]//Theory of Cryptography Conference. 2006: 265-284.
- [5] DWORK C, ROTH A. The algorithmic foundations of differential privacy[J]. *Found Trends Theor Comput SCI*, 2014, 9(3-4): 211-407.
- [6] ERLINGSSON Ú, PIHUR V, KOROLOVA A. Rappor: randomized aggregatable privacy-preserving ordinal response[C]//Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. 2014: 1054-1067.
- [7] BITTAU A, ERLINGSSON Ú, MANIATIS P, et al. Prochlo: strong privacy for analytics in the crowd[C]//Proceedings of the 26th Symposium on Operating Systems Principles. 2017: 441-459.
- [8] Differential Privacy Team, Apple. Learning with privacy at scale[S]. 2017.
- [9] DING B, KULKARNI J, YEKHANIN S. Collecting telemetry data privately[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 3574-3583.
- [10] JOHNSON N, NEAR J P, SONG D. Towards practical differential privacy for SQL queries[J]. *Proceedings of the VLDB Endowment*, 2018, 11(5): 526-539.
- [11] NISSIM K, WOOD A. Is privacy privacy[J]. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2018, 376(2128): 20170358.
- [12] KIFER D, MACHANAVAJJHALA A. No free lunch in data privacy[C]//Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data. 2011: 193-204.
- [13] COHEN A, NISSIM K. Towards formalizing the GDPR's notion of singling out[J]. *Proceedings of the National Academy of Sciences*, 2020, 117(15): 8344-8352.
- [14] OBERSKI D L, KREUTER F. Differential privacy and social science: an urgent puzzle[J]. *Harvard Data Science Review: HDSR*, 2020, 2(1): 1-21.
- [15] ROGAWAY P. The moral character of cryptographic work[J]. *Cryptology ePrint Archive*, 2015.
- [16] CUMMINGS R, KAPTCHUK G, REDMILES E M. I need a better description: an investigation into user expectations for differential privacy[C]//Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security. 2021: 3037-3052.
- [17] HOLZEL J. Differential privacy and the GDPR[J]. *Eur Data Prot L Rev*, 2019, 5: 184.
- [18] XIONG A, WU C, WANG T, et al. Using illustrations to communicate differential privacy trust models: an investigation of users' comprehension, perception, and data sharing decision[J]. *arXiv Preprint arXiv:2202.10014*, 2022.
- [19] XIONG A, WANG T, LI N, et al. Towards effective differential privacy communication for users' data sharing decision and comprehension[C]//2020 IEEE Symposium on Security and Privacy (SP). 2020: 392-410.
- [20] KIFER D, MACHANAVAJJHALA A. Pufferfish: a framework for mathematical privacy definitions[J]. *ACM Transactions on Database Systems (TODS)*, 2014, 39(1): 1-36.
- [21] MICHAEL C T, SEN S, DATTA A. SoK: differential privacy as a causal property[C]//2020 IEEE Symposium on Security and Privacy (SP). 2020: 354-371.
- [22] LIU C, CHAKRABORTY S, MITTAL P. Dependence makes you vulnerable: differential privacy under dependent tuples[C]//NDSS. 2016: 21-24.
- [23] ZHAO J, ZHANG J, POOR H V. Dependent differential privacy for correlated data[C]//2017 IEEE Globecom Workshops (GC Wkshps). 2017: 1-7.
- [24] ZHU T, XIONG P, LI G, et al. Correlated differential privacy: Hiding information in non-IID data set[J]. *IEEE Transactions on Information Forensics and Security*, 2014, 10(2): 229-242.
- [25] YANG B, SATO I, NAKAGAWA H. Bayesian differential privacy on correlated data[C]//Proceedings of the 2015 ACM SIGMOD international conference on Management of Data. 2015: 747-762.
- [26] ZHANG T, ZHU T, LIU R, et al. Correlated data in differential privacy: definition and analysis[J]. *Concurrency and Computation: Practice and Experience*, 2020.
- [27] LI Y, REN X, YANG S, et al. Impact of prior knowledge and data correlation on privacy leakage: a unified analysis[J]. *IEEE Transactions on Information Forensics and Security*, 2019, 14(9): 2017.

2342-2357.

- [28] CHEN R, FUNG B, YU P S, et al. Correlated network data publication via differential privacy[J]. The VLDB Journal, 2014, 23(4): 653-676.
- [29] SWEENEY L. k-anonymity: a model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570.
- [30] DALENIUS T. Towards a methodology for statistical disclosure control[J]. Statistik Tidskrift, 1977, 15(429-444).
- [31] DWORK C, NAOR M. On the difficulties of disclosure prevention in statistical databases or the case for differential privacy[J]. Journal of Privacy and Confidentiality, 2010, 2(1).
- [32] HIMMA K E, TAVANI H T. The handbook of information and computer ethics[M]. Hoboken: John Wiley & Sons, 2008.
- [33] GAVISON R. Privacy and the limits of law[J]. The Yale Law Journal, 1980, 89(3): 421-471.
- [34] BOK S. Secrets: on the ethics of concealment and revelation[M]. New York: Vintage, 1989.
- [35] TAVANI H T, MOOR J H. Privacy protection, control of information, and privacy-enhancing technologies[J]. ACM Sigcas Computers and Society, 2001, 31(1): 6-11.
- [36] WESTIN A F. Privacy and freedom[J]. Washington and Lee Law Review, 1968, 25(1): 166.
- [37] FRIED C. Privacy: a rational context[M]. New York: Computers, Ethics & Society. 1990: 51-63.
- [38] BERKMAN B A. The assault on privacy: computers, data banks, and dossiers[J]. Case Western Reserve Law Review, 1971, 22(4): 808.
- [39] RACHELS J. Why privacy is important[J]. Philosophy & Public Affairs, 1975: 323-333.
- [40] STAFF F T C. Protecting consumer privacy in an era of rapid change-a proposed framework for businesses and policymakers[J]. Journal of Privacy and Confidentiality, 2011, 3(1).
- [41] ABOWD J M. The US census Bureau adopts differential privacy[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018: 2867-2867.

[作者简介]



贾靖宇（1996-），男，天津人，南开大学博士生，主要研究方向为差分隐私。



谭畅（1996-），男，天津人，南开大学博士生，主要研究方向为差分隐私。



刘哲伟（2000-），男，吉林松原人，南开大学硕士生，主要研究方向为差分隐私。



李鑫豪（2000-），男，山东泰安人，南开大学硕士生，主要研究方向为差分隐私。



刘哲理（1978-），男，山东潍坊人，南开大学教授、博士生导师，主要研究方向为基于密码学的数据隐私保护、密文数据库、密文集合运算、差分隐私、人工智能安全。



张涛（1990-），男，四川苍溪人，中国工程物理研究院激光聚变研究中心高级工程师，主要研究方向为信息安全、数据融合分析与应用。