# 差分隐私模型中隐私参数 $\varepsilon$ 的选取研究

李 兰 杨 晨 王安福

(青岛理工大学信息与控制工程学院 山东 青岛 266000)

**摘 要** 差分隐私与传统的隐私保护方法不同,差分隐私可以对隐私保护强度进行量化分析,正是由于这一特点,使得差分隐私在数据发布、数据挖掘等方面得到了广泛的研究和应用。隐私预算因子 $\varepsilon$ 是影响隐私保护强度的重要因素之一,如何选取一个合理的 $\varepsilon$ 值,使数据的可用性达到最大化,并能够定量分析出隐私保护强度是亟待解决的一个问题。因此,通过分析满足 Laplace 分布噪音的概率密度函数与分布函数之间的关系,得到在噪音选取时,噪音可能落在的3种区间,从而建立隐私参数 $\varepsilon$ 与落点概率之间的数学关系表达式,并利用函数图像模型对参数 $\varepsilon$ 的选取计算式进行定量分析,最后结合攻击概率对隐私参数 $\varepsilon$ 的取值上界进行了探讨。

**关键词** 差分隐私,预算因子 $\varepsilon$,隐私保护,噪音干扰

**中图法分类号** TP309 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.08.033

## Study on Selection of Privacy Parameters $\varepsilon$ in Differential Privacy Model

LI Lan YANG Chen WANG An-fu

(School of Information and Control Engineering, Qingdao University of Technology, Qingdao, Shandong 266000, China)

**Abstract** Differential privacy is different from the traditional privacy protection methods. Differential privacy can quantify the privacy protection intensity. Because of this feature, differential privacy is widely studied and applied in data publishing and data mining. The privacy budget factor $\varepsilon$ is one of the important factors affecting the privacy protection intensity. How to choose a reasonable $\varepsilon$ value to maximize the availability of data and quantitatively analyze the privacy protection intensity is an urgent problem to be solved. Therefore, by analyzing the relationship between the probability density function and the distribution function satisfying the Laplace distributed, three kinds of noises in different range were choosen, so as to establish privacy parameter probability mathematical relational expression between epsilon and placement. And the function of image model was used to quantificationally analyze the selection formula of the parameter $\varepsilon$. Finally, the upper bound of privacy parameter epsilon was discussed combining with the attack probability.

**Keywords** Differential privacy, Budgets factor $\varepsilon$, Privacy preserving, Noise interference

## 1 引言

早在19世纪80年代,数据库隐私保护在国外就已经被统计学学者广泛研究和讨论,1977年,Dalenius提出了一种数据库语义安全的概念[1],即如果用户没有数据库内部访问权限,任何个人信息都不可能被获取,该想法的提出对后续研究者进行隐私保护的研究具有指导性意义。随着大数据时代的不断发展,我们使用过的各种服务、产品都成为与我们相关的数据源。2006年,Netfilx和AOL的隐私泄露事件令各大公司开始注重保护用户的隐私数据。研究者发现,公布一个数据集时,仅仅移除其中的姓名、身份识别码等敏感属性是没有用的,攻击者仍可以根据已有的知识背景推断出被移除的信息类别和属性。因此,2006年,微软的Dwork提出了一种隐私保护模型,即差分隐私(Differential Privacy)[2-6],它通过在发布的数据中添加适当的干扰噪音,使得攻击者即便已经掌握了除某一条信息以外的其他信息,仍然无法推测出被移除信息的属性内容。与传统的隐私保护技术[7-11]相比,$\varepsilon$-差分隐私的优点在于不需要特殊的攻击假设和背景知识,同时由于其定义建立在严格的数学统计模型上,可以给出严格的定量化分析来表示隐私披露风险,以最大化查询的准确率,最小化隐私泄漏的风险。因此,这是一种从数据源头彻底杜绝隐私信息泄露的可能性的方法。

实现差分隐私保护主要考虑3个方面[12-13]:1)满足差分隐私算法的可用性;2)满足差分隐私算法的效率;3)添加噪音所带来的误差对数据可用性的影响。

差分隐私的特殊之处在于通过改变隐私保护预算因子 $\varepsilon$ 的取值,可以控制发布数据的隐私程度。但 $\varepsilon$ 取值的合理性仍然是差分隐私面临的问题和挑战[14-16]。尽管差分隐私在

到稿日期:2018-06-12 返修日期:2018-11-25 本文受国家自然科学基金(61173181),国家自然科学基金(61772295)资助。

李 兰(1963-),女,硕士,教授,CCF会员,主要研究方向为数据挖掘、模式识别,E-mail:562474785@qq.com(通信作者);杨 晨(1992-),男,硕士生,主要研究方向为隐私挖掘、智能信息处理与模式识别;王安福(1994-),男,硕士生,主要研究方向为隐私挖掘、模式识别。

许多研究中被广泛应用,但隐私参数 ε 的选取大多是任意的,或者根据研究者的经验直接给定(例如给定 ε 为 0.01,0.1 等数值),很少有文献在不同的背景下对 ε 的取值进行探讨和研究。而实际上,对 ε 取值的研究不仅体现发布数据的隐私保护程度上,还体现在了可接受的数据披露风险的范围。

针对隐私参数 ε 的取值,本文提出了 3 种不同隐私级别下的噪音选取区间;同时,结合函数图像模型对不同假设区间内的 ε 选取计算式进行定量分析,并得出了不同假设条件下的 ε 取值上界的表达式。

## 2 相关研究

差分隐私作为一种严格的数学统计模型,在为数据发布的安全性和准确性提供保障的同时也引起了研究者的广泛关注。其研究方向和应用主要体现在面向数据挖掘和数据发布的差分隐私保护过程、控制隐私预算 ε 的合理性,以及保证其发布数据的准确性。

对基于差分隐私的数据发布研究问题,文献[2]最早提出利用差分隐私模型,在已经划分出的集合内添加噪音并以发布直方图的方式保护发布数据的隐私,但发布的数据的可用性会随着查询区间的增大而降低。鉴于此,Xiao 等[16]提出了一种新的解决方案——Privelet,该方案利用小波变换发布的直方图能够比较准确地响应较长范围的查询。在数据挖掘方面,文献[17]提出了一种满足 ε-差分隐私的 top-$k$ 频繁模式挖掘算法 DP-topkP。另外,Li 等[18]证明了在一定条件下,$k$-匿名可以满足(ε,λ)-差分隐私,同时将“安全 $k$ -匿名”通过 $\beta$-Sampling+Data-indepengent_Generalization +$k$-Suppression $(k,\beta)$-SDGS 变换,以提供差分隐私保护。

无论是差分隐私下的数据发布还是数据挖掘,都离不开选择一个合适的隐私预算因子 ε,ε 的取值直接影响了隐私保护强度。文献[19]定义了对同一个数据集 $D$,由 $i$ 个算法 $M_1,M_2,\cdots,M_{i-1},M_i$ 构成的 $M_i(X)$ 提供 $(\sum_i \varepsilon)$-差分隐私保护。文献[20]假设攻击者已知发布数据集的背景知识,那么以攻击成功概率 p 为参数,对于任意可能的数据集 $D'=D-1$,攻击者保留一个元组$\langle w,\alpha,\beta\rangle$,且对于每一个 $D'=w$,攻击者都可以通过一个查询响应得到返回值 $\alpha$ 和 $\beta$,其中 $\alpha$ 表示前验概率,$\beta$ 表示后验概率。假设前验概率满足均匀分布,当攻击概率 $p$ 小于后验概率 $\beta$ 时,其给出了隐私因子 ε 的选取上界 $\varepsilon\leqslant\frac{\Delta f}{\Delta v}\ln\frac{(n-1)\rho}{1-\rho}$。但是在一般情况下,前验概率并不满足平均分布,而且当 $D'$ 的数量很大时,隐私参数 ε 的值也会增大。文献[21]在文献[20]的基础上,提出一个新的攻击算法,在满足 Laplace 分布的条件下,给定一个长度为 $L$ 的查询区间,证明了 $q(D)+x$ 落在 $(-\infty,q(D)+\mu+L)$ 的概率为 $1-\frac{1}{2}e^{-\frac{L}{\lambda}}$,通过整理给出了 ε 的取值上界为:$\varepsilon\leqslant\frac{\ln 2(1-\rho)\Delta f}{L}$。根据文献[20,23],文献[22]在不依赖前验概率和已知数据表 $D'$ 全部可能值的前提下提出了一种新的 ε 设置方法 LPBDP。文献[24]首次提出一个全面模型,该模型可用来选择参数 ε 和 λ。其不仅仅提出并证明了 ε 的取值上界,还在不同的约束条件下提出了 ε 的取值下界。文献[25]定义了两个参数置信区间 $w$ 和置信水平 $p$,对于不同级别的差分隐私,当置信水平 $p$ 一定时,隐私保护强度会随着置信区间的增大而增大。其定义了对隐私保护的数学描述 $P[\hat{c}-wc<c<\hat{c}+wc]=p$,并通过对该描述求解逆函数得到满足 $p=1-e^{-\lambda wc}$ 和 $\varepsilon=-\frac{\ln(1-p)}{wc}$ 的数学表达式。

## 3 差分隐私的相关概念

### 3.1 ε-差分隐私模型

**定义 1**[2](相邻数据集) 给定属性结构相同的数据集 $D_1$ 和数据集 $D_2$,当两者记录相差的条数存在且为 1 时,称数据集 $D_1$ 和数据集 $D_2$ 为相邻数据集。

表 1 所列为相差一条记录的数据集 $D_1$ 和数据集 $D_2$,数据集 $D_2$ 除 Alice 的信息被完全删除外,其余记录信息与数据集 $D_1$ 完全相同,这两张表为相邻数据表。

表 1 数据集 $D_1$ 和数据集 $D_2$

Table 1 Dataset $D_1$ and dataset $D_2$

(a)数据集 $D_1$

| Name | Has cancer |
|---|---|
| Bob | 0 |
| Jim | 1 |
| Dick | 0 |
| Alice | 1 |

(b)数据集 $D_2$

| Name | Has cancer |
|---|---|
| Bob | 0 |
| Jim | 1 |
| Dick | 0 |

**定义 2**[2](ε-差分隐私) 相邻数据集 $D_1$ 和 $D_2$ 之间至多相差一条数据记录。给定随机算法 $K$,$Range(K)$ 表示随机算法 $K$ 的取值范围,$Pr[*]$ 表示数据集加上同一随机噪音之后查询结果为 $S$ 的概率。在 $D_1,D_2$ 上的输出结果 $S\subseteq Range(K)$ 均满足:

$$Pr[K(D_1)=S]\leqslant e^{\varepsilon}\times Pr[K(D_2)=S] \tag{1}$$

其中,ε 为隐私保护预算因子[26],用于衡量隐私保护的强度。相比于传统的 $k$-anonymity,ε-差分隐私给出了一个用户隐私不被泄露的概率上界,从数学概率的角度来看,ε 的取值决定了查询函数作用在发布数据表和原数据表的概率密度函数的相似度,当 ε 的取值趋近于 0,两种概率密度函数的相似度最高,即 $D_1$ 和 $D_2$ 的相似程度达到最高,隐私保护的程度最好。因此,ε 的取值越小,对于数据发布者越理想。

### 3.2 差分隐私的实现机制

**定义 3**[26](敏感度) 对于一个映射函数 $f:D\rightarrow R^d$ 表示数据集 $D$ 到一个 $d$ 维空间的映射,则 $f$ 的敏感度为:

$$\Delta f=\max_{D_1,D_2}\| f(D_1)-f(D_2)\|_p \tag{2}$$

**定义 4**[27](Laplace 机制) 在差分隐私研究中,Laplace 机制是最早被提出的差分隐私方法,该机制通过向查询数值 $f(D)$ 中添加满足 Laplace 分布的随机变量 $x$ 得到 $f(D)+x$ 来实现。其中,随机变量 $x$ 的概率密度函数为:

$$p(x|\mu,\lambda)=\frac{1}{2\lambda}e^{-\frac{|x-\mu|}{\lambda}} \tag{3}$$

其中,$\mu$ 代表位置参数,一般情况下默认为 0;$\lambda>0$ 为尺度参数,满足:

$$Lap(\lambda)=\frac{\Delta f}{\varepsilon} \tag{4}$$

输出函数 $A(D)$满足:

$$A(D)=f(D)+(Lap_1\ \frac{\Delta f}{\varepsilon}\ \cdots\ Lap_d(\frac{\Delta f}{\varepsilon}))^{\mathrm{T}} \tag{5}$$

对式(3)进行积分,可以得概率密度函数的分布函数 $P(x)$为:

$$P(x)=\begin{cases}\frac{1}{2}e^{-\frac{\mu-x}{\lambda}}, & x<\mu\\ 1-\frac{1}{2}e^{-\frac{x-\mu}{\lambda}}, & x\geqslant\mu\end{cases} \tag{6}$$

## 4　$\varepsilon$ 的取值分析

本节在已有背景知识的基础上从噪音值 $x$ 的取值角度,结合满足 Laplace 分布的概率密度函数模型与分布函数模型,对噪音值 $x$ 落在区间范围的概率进行分析。式(3)是描述随机变量噪音 $x$ 的输出值在某个确定的取值点附近的可能性的函数表达式,其图像如图 1 所示。从图中可以看出,当参数 $\lambda$ 取相同值、$\mu$ 取不同值时,曲线的平滑度是一致的,即发布数据添加的噪音幅度和发布数据集被成功攻击的概率与参数 $\mu$ 的选取无关。参数 $\lambda$ 不仅影响添加噪声的幅度,也是衡量隐私保护强度的重要参数。$\lambda$ 的取值越大,添加的噪音的分布曲线越平滑,发布数据集被成功攻击的可能性越低,数据的隐私保护程度越理想。

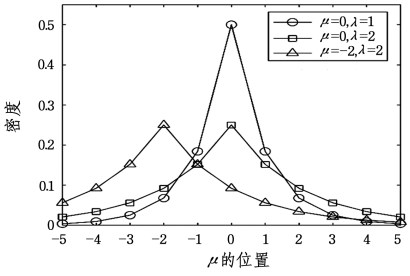


图 1　噪音 $x$ 的 Laplace 分布曲线

Fig. 1　Laplace distribution of noise $x$

根据概率密度函数的分布函数表达式(式(6)),可以建立一个满足 Lap($\mu$,$\lambda$|0,1)的累积分布函数,如图 2 所示。

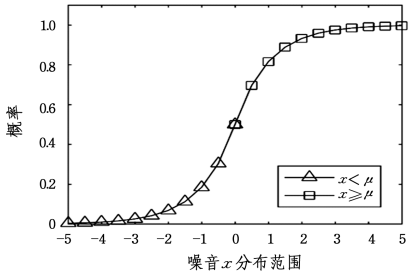


图 2　Lap($\mu$,$\lambda$|0,1)的概率分布图像

Fig. 2　Probability distribution of Lap($\mu$,$\lambda$ |0,1)

从图 2 中可以看到,累积分布函数 $P(x)$的值域为[0,1],该值域区间表示满足 Laplace 分布的随机变量噪音 $x$ 落在区间$(-\infty,\theta)(\theta\in(-\infty,+\infty))$之间的概率为[0,1]。因此,利用概率分布函数的特点,在生成满足 Laplace 分布的噪音时,首先生成一个分布在[0,1]区间上的随机值,然后将该随机值代入概率分布函数的反函数中,得到需要添加的噪声值 $x$。假设随机变量 $\xi\sim[0,1]$满足均匀分布,从图 2 所示的横纵坐标以及噪音 $x$ 的分布情况可以看出:

1)当 $\xi<\frac{1}{2}$时,函数图像对应概率分布函数中的 $P(x)=\frac{1}{2}e^{-\frac{x-\mu}{\lambda}}$,将 $\xi$ 代入 $P(x)$可以得到随机噪音 $x$ 满足的函数表达式:

$$x=\lambda\ln(2\xi)+\mu \tag{7}$$

2)当 $\xi\geqslant\frac{1}{2}$时,函数图像对应概率分布函数中的 $P(x)=1-\frac{1}{2}e^{-\frac{x-\mu}{\lambda}}$,将 $\xi$ 代入 $P(x)$可以得到随机噪音 $x$ 满足的函数表达式:

$$x=\mu-\lambda\ln(2(1-\xi)) \tag{8}$$

整理式(7)、式(8),可以得到添加噪音 $x$ 的逆累积分布函数:

$$x=\begin{cases}\lambda\ln(2\xi)+\mu, & \xi<\frac{1}{2}\\ \mu-\lambda\ln(2(1-\xi)), & \xi\geqslant\frac{1}{2}\end{cases} \tag{9}$$

从式(9)可以得到,若随机值 $\xi$ 的选择小于$\frac{1}{2}$,则 $x$ 所落区间的最大范围为$(-\infty,\mu)$。给定容错区间 $L$ 时,$f(D)+x$ 落在区间$(-\infty,f(D)+\mu-L)$的概率等于 $x$ 落在$(-\infty,\mu-L)$的概率,由此可以得出:

$$P(\mu-L)=\frac{1}{2}e^{-\frac{\mu-(\mu-L)}{\lambda}}=\frac{1}{2}e^{-\frac{L}{\lambda}}<\frac{1}{2}$$

图 3 为在区间$(-\infty,\mu-L)$内不同隐私参数 $\varepsilon$ 值对应噪音值落在该区间内的概率曲线图像。从图 3 中可以看出,随着隐私参数 $\varepsilon$ 的增大,$f(D)+x$ 落在给定区间内的概率变小。

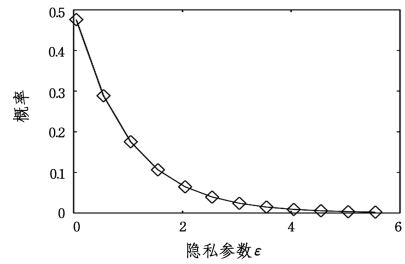


图 3　$(-\infty,\mu-L)$内 $\varepsilon$ 与 $P(\mu-L)$的关系

Fig. 3　Relationship between $\varepsilon$ and $P(\mu-L)$ in $(-\infty,\mu-L)$

假设存在一个小于$\frac{1}{2}$的攻击成功概率 $\eta$,若 $f(D)+x$ 落在区间$(-\infty,f(D)+\mu-L)$内,由于存在尺度参数 $\lambda=\frac{\Delta f}{\varepsilon}$,且隐私保护强度与 $\lambda$ 呈正相关,因此只有在$\frac{1}{2}>\frac{1}{2}e^{-\frac{L}{\lambda}}>\eta$时,发布数据的隐私保护强度较高。通过化简该不等关系可以得到 $\varepsilon$ 满足的不等式关系:

$$\varepsilon<-\frac{\Delta f\ln 2\eta}{L}$$

若随机值 $\xi$ 的选择大于或等于 $\frac{1}{2}$，则 $x$ 所落区间的范围是 $(\mu,+\infty)$，$f(D)+x$ 落在 $(-\infty,f(D)+\mu+L)$ 的概率等于 $x$ 落在 $(-\infty,\mu+L)$ 的概率，通过式(6)可以得到 $x$ 落入区间 $(-\infty,\mu+L)$ 内的概率为：

$$P(\mu+L)=1-\frac{1}{2}e^{-\frac{(\mu+L)-\mu}{\lambda}}=1-\frac{1}{2}e^{-\frac{L}{\lambda}}$$

如图 4 所示为在区间 $(-\infty,\mu+L)$ 内，随着 $\varepsilon$ 值的不断变大，$f(D)+x$ 落在给定区间内的概率也变大。

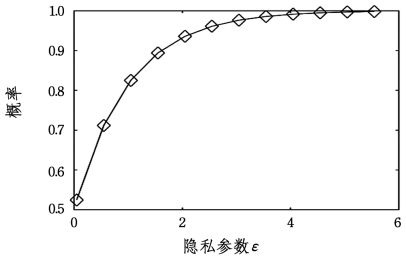


图 4 $(-\infty,\mu+L)$ 内 $\varepsilon$ 与 $P(\mu+L)$ 的关系

Fig. 4 Relationship between $\varepsilon$ and $P(\mu+L)$ in $(-\infty,\mu+L)$

假设存在一个大于 $\frac{1}{2}$ 的攻击成功概率 $\eta$，若 $f(D)+x$ 落在区间 $(-\infty,f(D)+\mu+L)$ 内，当 $\frac{1}{2}<\eta\leqslant 1-\frac{1}{2}e^{-\frac{L}{\lambda}}$ 时，攻击者很难得到查询值是否在查询区间之内，通过化简，可以得到与 $\varepsilon$ 有关的不等式关系：

$$\varepsilon\leqslant\frac{\Delta f\ln 2(1-\eta)}{L}$$

另外，从容错区间 $L$ 的角度看，假设存在区间 $[\mu-L,\mu+L]$，通过概率密度函数表达式(式(3))在区间 $[\mu-L,\mu+L]$ 上的积分，可以得到 $f(D)+x$ 落在区间 $[\mu-L,\mu+L]$ 的概率 $P(\gamma)$ 为：

$$\begin{aligned}P(\gamma)&=\int_{\mu-L}^{\mu+L}\frac{1}{2\lambda}e^{-\frac{|x-\mu|}{\lambda}}dx\\&=\int_{\mu-L}^{0}\frac{1}{2\lambda}e^{-\frac{|x-\mu|}{\lambda}}dx+\int_{0}^{\mu+L}\frac{1}{2\lambda}e^{-\frac{|x-\mu|}{\lambda}}dx\\&=e^{\frac{\mu-L}{\lambda}}\end{aligned}$$

图 5 所示为在区间 $[\mu-L,\mu+L]$ 内隐私参数 $\varepsilon$ 的不同取值与对应噪音值落在区间内的概率曲线图。从图中可以看出，在 $[\mu-L,\mu+L]$ 区间范围内，$x$ 落在区间 $[\mu-L,\mu+L]$ 的概率会随着 $\varepsilon$ 值的不断变大而减小，即 $f(D)+x$ 落在区间 $[\mu-L,\mu+L]$ 内的概率会随着 $\varepsilon$ 值的变大而变小。

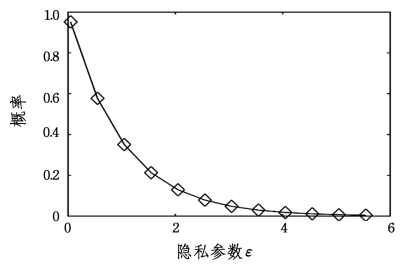


图 5 $[\mu-L,\mu+L]$ 内 $\varepsilon$ 与 $P(\gamma)$ 的关系

Fig. 5 Relationship between $\varepsilon$ and $P(\gamma)$ in $[\mu-L,\mu+L]$

因此，当攻击成功的概率 $\eta$ 小于 $f(D)+x$ 落在区间内的概率 $P(\gamma)$ 时，即 $\eta\leqslant e^{\frac{\mu-L}{\lambda}}$，此时的隐私保护程度较高。将不等式化简，可以得到与 $\varepsilon$ 有关的不等式关系：

$$\varepsilon\leqslant\frac{\Delta f\ln p}{\mu-L}$$

**结束语** 预算因子 $\varepsilon$ 是衡量隐私保护强度的重要因素，其不同的分配方案对隐私保护算法的误差有着较大的影响。本文在已有研究的基础上，针对噪音选取时的随机变量 $\xi$，考虑了两种 $\xi$ 的取值情况。同时，通过结合查询区间 $L$ 和攻击者攻击发布数据成功的概率 $\eta$，对 $\varepsilon$ 的取值上界进行了分析；另外，在给定查询区间内，推导出了 $\varepsilon$ 取值上界的函数关系。对 $\varepsilon$ 的研究不仅仅局限于在 Laplace 机制中选取一个合适的隐私参数值，在指数机制中选取一个合理的 $\varepsilon$ 和利用概率统计的参数估计方法计算出一个理想的参数值也是值得研究的方向。

## 参 考 文 献

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

[2] DWORK C. Differential privacy[C]// Proceedings of the 33rd International Colloquium on Automata, Languages and Programming. Berlin: Springer, 2006: 1-12.

[3] DWORK C. Differential privacy: A survey of results[C]// Proceedings of the 5th International Conference on Theory and Applications of Models of Computation. Berlin: Springer-Verlag, 2008: 1-19.

[4] DWORK C. Differential privacy and robust statistics[C]// Proceedings of the 41st Annual ACM Symposium on Theory of Computing. New York: ACM Press, 2009: 371-380.

[5] DWORK C, NARO M, REINGOLD O, et al. On the complexity of differentially private data release: efficient algorithms and hardness results[C]// Proceedings of the 41st Annual ACM Symposium on Theory of Computing. New York: ACM Press, 2009: 381-390.

[6] DWORK C. The differential privacy frontier[C]// Proceedings of the 6th Theory of Cryptography Conference on Theory of Cryptography. Berlin: Springer, 2009: 496-502.

[7] SWEENEY L. *k*-anonymity: a model for protecting privacy[J]. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002, 10(5): 557-570.

[8] MACHANAVAJJHALA A, GEHRKE J, KIFER D. *l*-diversity: privacy beyond *k*-anonymity [J]. ACM Transactions on Knowledge Discovery from Data, 2007, 1(1): 24-35.

[9] LI N, LI T, VENKATASUBRAMANIAN S. *t*-closeness: privacy beyond *k*-anonymity and *l*-diversity[C]// Proceedings of the IEEE International Conference on Data Engineering. Washington DC: IEEE Computer Society, 2007: 106-115.

[10] WONG C W, LI J, FU W C, et al. ($\alpha$-*k*)-anonymity: An enhanced *k*-anonymity model for privacy preserving data publishing [C]// Proceedings of the 12th ACM SIGKDD International

Conference on Knowledge Discovery and Data Mining. New York：ACM Press，2006：754-759.

[11] XIAO X，TAO Y. *m*-invariance：towards privacy preserving republication of dynamic datasets[C]// Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data. New York：ACM Press，2007：689-700.

[12] 吴英杰. 隐私保护数据：发布模型与算法[M]. 北京：清华大学出版社，2015.

[13] DWORK C，SMITH A. Differential privacy for statistics：What we know and what we want to learn [J]. Journal of Privacy and Confidentiality，2010，1(2)：135-154.

[14] XIONG P，ZHU T Q，WANG X F. A Survey on Differential Privacy and Applications[J]. Chinese Journal of Computers，2014，37(1)：101-102. (in Chinese)
熊平，朱天清，王晓峰. 差分隐私保护及其应用[J]. 计算机学报，2014，37(1)：101-102.

[15] ZHANG X J，MENG X F. Differential privacy in data publication and analysis[J]. Chinese Journal of Computers，2014，37(4)：927-949. (in Chinese)
张啸剑，孟小峰. 面向数据发布和分析的差分隐私保护[J]. 计算机学报，2014，37(4)：927-949.

[16] XIAO X，WANG G，GEHRKE J. Differential privacy via wavelet transforms[J]. IEEE Transon Knowledge and Data Engineering，2012，23(8)：1200-1214.

[17] ZHANG X J，WANG M，MENG X F. An accurate method for mining top-k frequent pattern under differential privacy[J]. Journal of Computer Research and Development，2014，51(1)：104-114. (in Chinese)
张啸剑，王淼，孟小峰. 差分隐私保护下一种精确挖掘 top-*k* 频繁模式方法[J]. 计算机研究与发展，2014，51(1)：104-114.

[18] LI N，QARDAJI W，SU D. Provably Private Data Anonymization：Or，*k*-anonymity meets differential privacy，CERIAS TR2010-24 [R]. West Lafayette：Center for Education and Research Information Assurance and Security，Purdue University，2010.

[19] McSHERRY F. Privacy integrated queries：An extensible platform for privacy-preserving data analysis [C]// Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. New York：ACM Press，2009：19-30.

[20] LEE J，CLIFTON C. How much is enough? Choosing ε for differential privacy[C]// Proceeding of the 14th International Conference on Information Security. Berlin：Springer，2011：325-340.

[21] HE X M，WANG X Y，CHEN H H，et al. Study on choosing the parameter ε in differential privacy[J]. Journal on Communications，2015，36(12)：124-130. (in Chinese)
何贤芒，王晓阳，陈华辉，等. 差分隐私保护参数 ε 的选取研究[J]. 通信学报，2015，36(12)：124-130.

[22] OUYANG J，XIAO Z H，LIU S P，et al. Heuristic privacy parameter setting strategy for differential privacy model[J/OL]. Application Research of Computers. http://www. arocmag. com/ article/02-2019-01-037. html. (in Chinese)
欧阳佳，肖政宏，刘少鹏，等. 差分隐私模型的启发式隐私参数设置策略[J/OL]. 计算机应用研究. http://www. arocmag. com/article/02-2019-01-037. html.

[23] EVFIMIEVSKI A，GEHRKE J，SRIKANT R. Limiting privacy breaches in privacy preserving data mining [C]// Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. New York：ACM Press，2003：211-222.

[24] HSU J，GABOARDI M，et al. Differential Privacy：An Economic Method for Choosing Epsilon[C]// Proceedings of the IEEE 27th Computer Security Foundations Symposium. 2014：398-410.

[25] NALDI M，ACQUISTO D G. Differential Privacy：An Estimation Theory-Based Method for Choosing Epsilon[J]. arXiv：1510.00917.

[26] HAEVERLEN A，PIERCE B C，NARAYA A. Differential privacy under fire [C]// Proceedings of the 20th USENIX Conference on Security. 2011：33-39.

[27] DWORK C，McSHERRY F，NISSIM K，et al. Calibrating Noise to Sensitivity in Private Data Analysis[C]// Proceedings of the 3rd Conference on Theory of Cryptography. Berlin：Springer，2006：265-268.