

基于置信度分析的差分隐私保护参数配置方法研究

李森有, 季新生, 游伟

(国家数字交换系统工程技术研究中心, 河南 郑州 450002)

摘要: 为了解决数据发布和分析过程中用户真实数据信息被披露的问题, 降低攻击者通过差分攻击和概率推理攻击获取真实结果的概率, 提出了一种基于置信度分析的差分隐私保护参数配置方法。在攻击者概率推理攻击模型下对攻击者置信度进行分析, 使之不高于根据数据隐私属性所设置的隐私概率阈值。所提出的方法能够针对不同查询用户查询权限的差异配置更加合理的隐私保护参数, 避免了隐私披露的风险。实验分析表明, 所提出的方法根据查询权限、噪声分布特性以及数据隐私属性分析攻击者置信度与隐私保护参数的对应关系, 并据此推导出隐私保护参数的配置公式, 从而在不违背隐私保护概率阈值的情况下配置合适的 ϵ 参数。

关键词: 差分隐私; 置信度分析; 概率推理攻击模型; 隐私保护

中图分类号: TP309.2

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2019036

Research on differential privacy protection parameter configuration method based on confidence level

LI Senyou, JI Xinsheng, YOU Wei

National Digital Switching System Engineering & Technological Research Center, Zhengzhou 450002, China

Abstract: In order to solve the problem that the user's real data information is disclosed during the data release and analysis process, and reduce the probability of an attacker gaining real results through differential attacks and probabilistic inference attacks, a differential privacy protection parameter configuration method based on confidence level is proposed. Analysis of attacker confidence under attacker probabilistic inference attack model and make it no higher than the privacy probability threshold set according to the data privacy attribute. The proposed method can configure more reasonable privacy protection parameters for different query privilege of query users, and avoids the

收稿日期: 2019-03-20; 修回日期: 2019-06-06

通信作者: 李森有, lisenyou1993@163.com

基金项目: 国家自然科学基金创新研究群体资助项目 (No.61521003, No.61801515); 国家重点研发计划基金资助项目 (No.2016YFB0801605)

Foundation Items: The National Natural Science Foundation for Creative Research Groups of China (No.61521003, No.61801515), The National Key R&D Program of China (No.2016YFB0801605)

论文引用格式: 李森有, 季新生, 游伟. 基于置信度分析的差分隐私保护参数配置方法研究[J]. 网络与信息安全学报, 2019, 5(4): 29-39.

LI S Y, JI X S, YOU W. Research on differential privacy protection parameter configuration method based on confidence level[J]. Chinese Journal of Network and Information Security, 2019, 5(4): 29-39.

risk of privacy disclosure. The experimental analysis shows that the proposed method analyzes the correspondence between attacker confidence level and privacy protection parameters based on query privilege, noise distribution characteristics and data privacy attributes, and derives the configuration formula of privacy protection parameters, which configure the appropriate parameters without violating the privacy protection probability threshold.

Key words: differential privacy, confidence level, probability inference attack model, privacy protection

1 引言

近年来,随着信息技术的迅猛发展,用户数据爆炸式增长^[1]。通过数据挖掘和信息收集提取的个人信息成为各类研究机构、组织和政府部门等研究和决策的宝贵资源。海量用户数据的分析和使用在给人们生活带来便利的同时也给用户隐私保护带来巨大的威胁。20 世纪以来,如何在数据发布过程中保护个人隐私数据以抵御统计披露攻击成为隐私保护领域的研究重点^[2]。早期的基于数据失真、基于数据加密以及基于限制发布的隐私保护技术通过 k -anonymity、 l -diversity 以及 t -closeness 等方法^[3-5]泛化敏感信息的标志属性在一定程度上能够抵抗部分攻击,但是对于一些新的攻击模型(如背景知识攻击、相似性攻击和倾斜攻击等)仍会面临隐私数据泄露的威胁。

差分隐私^[6-8]是 Dwork 针对统计数据库的隐私泄露问题提出的一种新的基于严格数学背景的隐私保护机制,通过使用输出扰动技术对查询输出进行添加随机噪声处理,使单个记录在数据集中或者不在数据集中,对计算结果的影响微乎其微。即使在攻击者拥有最大背景知识的条件下,也能够保证攻击者无法通过观察计算结果获取准确的个体信息。同时为隐私保护度量提供了量化评估方法,使不同隐私保护参数下的隐私保护水平具有可比性^[9]。

但目前差分隐私保护研究工作主要集中在提高差分隐私数据发布中的隐私保护和数据效用^[10-12],而很少对差分隐私具体实践中隐私保护参数的配置方法进行严格的数学推理和研究。虽

然差分隐私保护参数 ϵ 可以用于量化敏感数据发布所带来的隐私披露风险,但文献[13]指出在具体的实践中如何根据不同的参数和用户需求选择合适的 ϵ 值还没有严格的方法,即说现有的隐私参数选择并不是根据需要自由选择的。在具体的实践中,隐私保护参数 ϵ 值的配置需要考虑数据集大小,查询函数敏感度,隐私保护概率阈值等。例如,文献[14]以差分隐私保护下的平均值和中间值发布为例,提出了一种攻击模型,根据攻击者接收相应的先验和后验概率,通过数学推导分析隐私保护参数 ϵ 的选取上界以及与其他参数的数学对应关系。文献[15]提出了一个新的差分隐私保护攻击模型,假设攻击者可以根据查询语句的返回值回答攻击对象是否在查询数据集合中。通过分析推导出攻击者成功概率与查询函数敏感度、容错区间以及隐私保护参数 ϵ 选取的数学表达式,进而优化隐私保护参数 ϵ 选取,实现更好的差分隐私保护。文献[16]则在文献[15]的基础上增加了攻击者连续多次攻击的分析,提出一种重复攻击下实现差分隐私保护技术安全性的攻击模型。通过对重复攻击的结果与噪声分布函数的特点进行分析,获取噪声落入容错区间的概率,结合排列组合方法,推断出攻击者攻击成功的概率,进而得到参数 ϵ 的选取范围。

本文参考文献[15]对差分隐私保护机制中攻击者推断成功概率研究,对差分隐私保护中攻击者置信度进行分析,设置隐私保护概率阈值以实现隐私披露概率的控制,提出了一种基于置信度分析的差分隐私保护参数配置方法。首先,根据攻击者概率推理攻击模型对不同置信区间、噪

声分布的位置参数以及尺度参数下的攻击者置信度进行分析；然后，根据所查询数据的隐私保护属性设置隐私保护概率阈值，在配置隐私保护参数时保证攻击者置信度不高于所设置的隐私保护概率阈值；最后，针对不同查询用户查询权限的差异，研究用户查询权限对隐私保护参数配置的影响，从而配置更加合理的隐私保护参数。同时在差分隐私保护模型中引入可用性评估模块，在保护隐私的同时对数据的可用性进行分析，实现了可控的敏感数据隐私性和可用性保护，大大降低了隐私泄露的危险。

2 基本定义

本节介绍一些差分隐私的定义和相关概念。

定义 1^[17] 差分隐私

设有定义域为 D ，值域为 R 的随机函数 $M: D \rightarrow R$ ， R 为 M 所有可能的输出构成的集合， $Sm \subseteq \text{Range}[M]$ 。对于任意两个差别至多为一个记录的相邻数据集 $D1$ 和 $D2$ ，若算法 M 满足式(1)，则称算法 M 提供差分隐私保护。

$\Pr[M(D1) \in Sm] \leq \exp(\varepsilon) \times \Pr[M(D2) \in Sm]$ (1)

其中， $\Pr[M]$ 表示隐私被披露的概率， ε 称为隐私保护预算参数，用来控制算法 M 在两个邻近数据集上输出概率的比值，反映了 M 可以提供的差分隐私保证的强度。 ε 取值越小，隐私保护程度越高。

定义 2 函数敏感度^[18]

对于给定的查询函数 $f: D \rightarrow Rd$ ，其敏感度为

$$\Delta f = \max_{(D1, D2)} \|f(D1) - f(D2)\|_1 = \max_{(D1, D2)} \sum_{i=1}^k |f(D1)_i - f(D2)_i| \quad (2)$$

其中， $\|f(D1) - f(D2)\|_1$ 表示相邻数据集 $D1$ 和 $D2$ 函数输出值的一阶范数距离。

函数敏感度是决定在查询函数的返回值中添加

加噪声量大小的重要参数，仅和查询函数本身性质相关，是指在数据集中删除或添加任意一条记录后对输出结果产生的最大影响。

定理 1 (Laplace 机制^[18]): 对于给定数据集 D ，设有查询函数 $f: D \rightarrow Rd$ ，其敏感度为 Δf ，如果满足下列公式，则表达式 $M(D)$ 满足 ε -差分隐私保护。

$$M(D) = f(D) + \left(\text{Laplace}\left(\frac{\Delta f}{\varepsilon}\right) \right)^d \quad (3)$$

其中， $\text{Laplace}\left(\frac{\Delta f}{\varepsilon}\right)$ 是随机噪声，服从尺度参数为 $\left(\frac{\Delta f}{\varepsilon}\right)$ 的 Laplace 分布，其概率密度函数为

$$P(x) = \frac{\varepsilon}{2\Delta f} \exp\left(-\frac{\varepsilon|x|}{\Delta f}\right)。$$

3 概率推理攻击模型

3.1 应用差分隐私

在差分隐私保护机制中，在输出结果中添加随机噪声从而扰动真实结果，在数据失真的同时保证其统计特征。具体来说，在发布统计信息的数据库查询中， ε -差分隐私保护机制对查询函数 f 给出扰动响应 $f(x) + Y$ 而不是真实结果 $f(x)$ 。其中，所添加的随机噪声 Y 的幅度决定了隐私保护的强度，噪声幅度越大隐私保护效果越好。

3.2 概率推理攻击模型

如上文所述，在差分隐私保护机制中，攻击者获得的查询返回值是在真实值基础上添加噪声后生成的，噪声的分布直接影响攻击者获得真实信息的概率。因此本节描述一种攻击模型，假设攻击者拥有最大背景知识。当多个查询者对统计数据库提出同一查询请求时，数据集管理者对这些查询者不加以区分而将添加噪声后的结果 $f(x) + Y$ 返回给所有查询者。对于恶意攻击者而言，其具有较低的查询权限但却获取了与其他权

限和信誉较高查询用户相同隐私保护程度的数据。攻击者为了从查询返回结果中推断出真实结果,在进行差分攻击获取差异化查询结果后基于置信区间分析进行概率推理攻击,差分攻击和概率推理攻击定义如下。

定义 3 差分攻击

差分攻击是一种当统计数据库中插入或删除一个记录时由查询结果的差异来推断记录值的攻击。例如攻击者知道某个公司有 20 个职员,平均工资为 5 000 元,攻击者通过查询得到在排除职员 A 后该公司员工的平均工资为 4 900 元,则攻击者可以得到职员 A 的工资为 6 900 元 ($20 \times 5000 - 19 \times 4900$)。在这种情况下,差分攻击可以表达如下

$$x = N \cdot Avg_{\text{before}} - (N-1)Avg_{\text{after}} \quad (4)$$

定义 4 基于置信区间分析的概率推理攻击

概率推理攻击意味着攻击者能够从查询返回结果中通过使用噪声分布规律结合其置信区间设定推断查询的真实结果。概率推理攻击过程如下。

首先,攻击者获取的查询返回结果中的噪声都是服从某种特定分布的,并且噪声分布特性受隐私保护参数 ε 决定。如果隐私保护参数 ε 取值较大,则查询返回结果会集中在某一个值附近,攻击者通过差分攻击推断所添加的噪声量大小。对于某些查询函数,如果噪声 x 集中分布在 $[-\alpha, \alpha]$ ($\alpha > 0$),那么攻击者能够大概率的推理出真实值 $f(x)$,从而分析出某个具体记录在或者不在数据集中。本文把 $[-\alpha, \alpha]$ 称为攻击者概率推理攻击的置信区间,对应的置信度为 $conf(x)$ 。置信度表示对置信区间估计的把握程度,较宽的置信区间代表把握程度越大,但相应降低了估计的准确程度。

Laplace 概率密度曲线如图 1 所示,服从 Laplace 分布的随机变量 x 的概率密度函数为

$$f(x|\mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right) = \frac{1}{2b} \begin{cases} \exp\left(-\frac{\mu-x}{b}\right), & x < \mu \\ \exp\left(-\frac{x-\mu}{b}\right), & x \geq \mu \end{cases} \quad (5)$$

其中,位置参数 μ 和尺度参数 $b = \frac{\Delta f}{\varepsilon}$ 直接影响数据的分布情况。当尺度参数 b 取值较小时,服从 Laplace 分布的数据大多集中在位置参数 μ 附近。当攻击者对数据库进行计数查询请求时,其使用差分攻击和概率推理攻击得到噪声 x 集中分布在置信区间 $[-0.5, 0.5]$ 时(如图 1 所示各分布曲线阴影区域),攻击者可以很容易得出所查询问题的真实值,进而分析得出所攻击对象的隐私信息。在这种情况下,攻击者的置信区间选取以及噪声分布情况反映了攻击者获取某个具体记录隐私信息的难易程度。

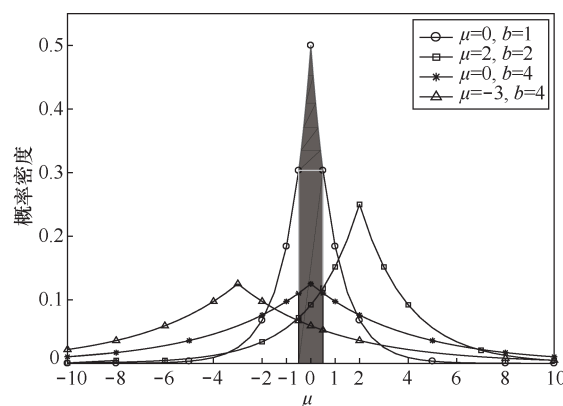


图 1 置信区间 $[-0.5, 0.5]$ 的 Laplace 分布

4 基于置信度分析的参数配置方法

为了防止差分隐私概率推理攻击,本文提出了一种基于置信度分析的差分隐私保护参数配置方法,在保持数据效用的条件下,在所提出的差分隐私保护框架下,通过配置合理的隐私保护参数添加适量的噪声达到隐私保护的目的。

表1 符号列表

符号	含义
D	数据库 D
E	隐私保护参数
$Lap()$	Laplace 分布函数
B	Laplace 尺度参数
Δf	查询函数敏感度
$[-\alpha, \alpha]$	置信区间
$conf(x)$	置信度
T_{prob}	隐私保护概率阈值

4.1 差分隐私保护框架

如图2所示,本文提出一种基于置信度分析的差分隐私保护参数配置方法。为了防止攻击者通过差分攻击和概率推理攻击,对决定所添加噪声幅度的隐私保护参数的配置方法进行研究。首先分析攻击者的查询请求及其查询权限,根据这个结果可以做出一个概率推理模型分析攻击者获取真实结果的把握,即置信度(由置信区间决定)。然后,根据数据的隐私保护属性确定隐私保护概率阈值,确保攻击者获取真实结果的置信度不超过所设定的概率阈值,从而推导出差分隐私保护参数配置的数学表达式。最后,当数据查询者提出一个查询请求时,它会得到根据导出参数添加

噪声的不精确信息。同时,在保护模型中添加可用性评估模块,衡量差分隐私处理后的数据可用性,实现可控的敏感数据隐私性和可用性保护,能够有效防止恶意攻击者通过信息查询获取用户隐私信息,在降低隐私泄露的前提下大大提高发布数据的有效使用率。

4.2 基于置信度的差分隐私保护方法分析

4.2.1 隐私性分析

数据的隐私保护效果最重要的是通过对数据集处理前后数据分布的差异来侧面反映,差异程度直接影响攻击者获取真实数据的概率。差分隐私保护技术中,在数据查询的真实结果中添加符合 Laplace 分布的噪声保证 ε -差分隐私。

$$\begin{aligned} \frac{\Pr_{D_1}(x)}{\Pr_{D_2}(x)} &= \prod_{i=1}^k \left(\frac{\exp\left(-\frac{\varepsilon|f(D_1)_i - x_i|}{\Delta f}\right)}{\exp\left(-\frac{\varepsilon|f(D_2)_i - x_i|}{\Delta f}\right)} \right) = \\ &= \prod_{i=1}^k \exp\left(\frac{\varepsilon(|f(D_2)_i - x_i| - |f(D_1)_i - x_i|)}{\Delta f}\right) \leq \\ &= \prod_{i=1}^k \exp\left(\frac{\varepsilon|f(D_1)_i - f(D_2)_i|}{\Delta f}\right) = \\ &= \exp\left(\frac{\varepsilon\|f(D_1) - f(D_2)\|_1}{\Delta f}\right) \leq \exp(\varepsilon) \end{aligned} \quad (6)$$

由此得知对于 ε -差分隐私保护, ε 越小,输

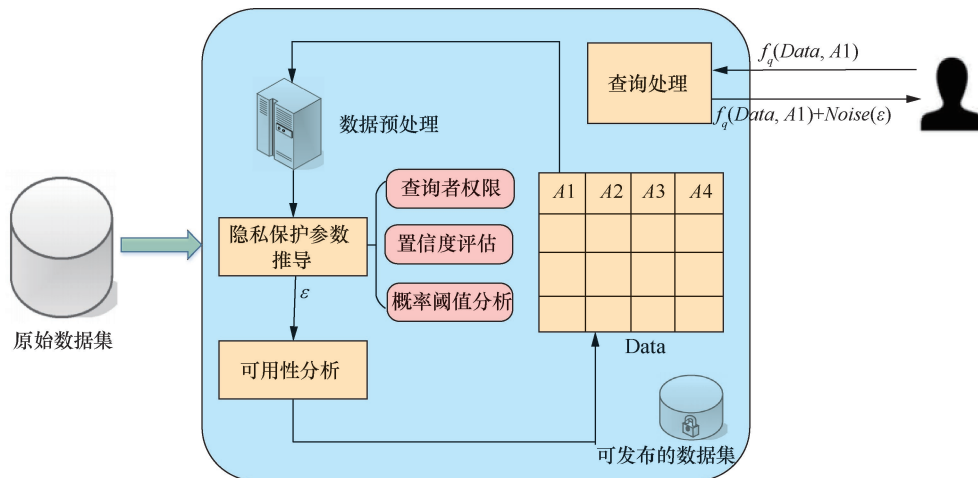


图2 基于置信度分析的噪声参数配置方法

出值差异越小。当 ε 取 0 时, 数据集处理前后以相同的概率输出相同的值, 这时隐私保护程度最高, 但同时数据的可用性最低。

4.2.2 可用性分析

在本文的研究中, 笔者在差分隐私保护框架中添加可用性分析模块来衡量差分隐私保护算法中所添加噪声对数据可用性的影响。从理论角度, 通常采用 (α, β) -useful^[7,19] 技术。

(α, β) -useful: 对于差分隐私算法 M , 设 Q 是一组查询函数, X 是数据域。如果对于任何 n 项数据库 x , 对于 Q 和 X , 若满足式(7), 则算法 M 满足 (α, β) -useful。

$$\Pr[|M(D1) - M(D2)| \leq \alpha] > 1 - \beta \quad (7)$$

在具体的实际应用中, 常根据不同的使用场景选择不同的差度量方法, 如相对误差、绝对误差、欧拉函数以及 F-measure 等。根据需要可以选择平均相对误差作为数据库处理前后数据可用性分析的标准。

$$\text{ARE}(Q(D)) = \sum_{i=1}^n \frac{1}{n} \frac{|Q_i(D) - Q_i(D')|}{\max(Q(D), s)} \quad (8)$$

其中, $Q_i(D)$ 为数据集第 i 次查询的真实输出结果, $Q_i(D')$ 为相对应的隐私保护处理后的查询输出结果。 s 是当 $Q_i(D)$ 出现特别小的情况时的数值下限。

可用性分析模块可以判断差分隐私保护查询返回结果的数据的隐私性和可用性是否满足用户的需求, 向隐私保护模型反馈误差信息。当查询处理结果的数据可用性低于用户隐私要求或隐私保护程度不满足隐私保护概况阈值设定时, 则重新进行添加噪声处理, 直至得到满足要求的结果。

4.2.3 置信度分析

为了解决数据发布查询中攻击者的概率推理攻击问题, 本文提出利用攻击者置信度分析来衡量隐私泄露概率。3.2 节中给出提出置信区间和置

信度的概念, 以图 1 中 Laplace 分布为例, 很直观地表示了置信度和置信区间的关系。对于相同的置信区间, 区间内概率密度越大, 相应的置信度越大, 攻击者获取真实结果的概率越大。本文主要以 Laplace 噪声分布为例进行研究, 根据查询函数特性和噪声分布特性计算置信度与置信区间、函数敏感度和查询权限之间的数学对应关系。从 Laplace 定义出发, 随机噪声 x 位于置信区间 $[-\alpha, \alpha]$ 的概率可由 $F(\alpha) - F(-\alpha)$ 得到。其中, Laplace 累积分布函数 $F(x)$ 定义为

$$F(x) = \int_{-\infty}^x f(u) du = \begin{cases} 1 - \frac{1}{2} e^{-\frac{x-\mu}{b}}, & x \geq \mu \\ \frac{1}{2} e^{\frac{x-\mu}{b}}, & x < \mu \end{cases} = \frac{1}{2} \left[1 + \text{sgn}(x - \mu) \left(1 - e^{-\frac{|x-\mu|}{b}} \right) \right] \quad (9)$$

由此可得攻击者置信度 $\text{conf}(x)$ 的数学表达式为

$$\text{conf}(x) = F(\alpha) - F(-\alpha) = \int_{-\alpha}^{\alpha} f(u) du = \begin{cases} \frac{1}{2} \left(\exp\left(\frac{\mu+\alpha}{b}\right) - \exp\left(\frac{\mu-\alpha}{b}\right) \right), & \mu \leq -\alpha \leq \alpha \\ 1 - \frac{1}{2} \left(\exp\left(\frac{\mu-\alpha}{b}\right) + \exp\left(\frac{-\alpha-\mu}{b}\right) \right), & -\alpha \leq \mu \leq \alpha \\ \frac{1}{2} \left(\exp\left(\frac{\alpha-\mu}{b}\right) - \exp\left(\frac{-\alpha-\mu}{b}\right) \right), & -\alpha \leq \alpha \leq \mu \end{cases} \quad (10)$$

通过这个数学理论分析, 可以选取合适的隐私保护参数 ε , 在查询结果中添加满足差分隐私保护要求的噪声, 达到防止攻击者概率推理攻击的目的。

4.3 隐私保护参数配置

本文提出的基于置信度分析的隐私保护参数配置方法旨在确定合适的噪声参数 ε , 从而抵御攻击者概率推理攻击。为了实现这一目的, 在前文中计算出了攻击者置信度的数学表达式, 该公

式可以由 Laplace 分布的尺度参数 b 表示。当攻击者置信度满足隐私概率阈值设定时，可以得到合适的尺度参数 b 的值。在提出的方法中，隐私概率阈值 $T_{\text{prob}} \in (0,1)$ 由所查询的数据隐私保护属性确定，意味着攻击者的概率推理攻击不会超过隐私保护界限。

通过以上的分析，为了满足隐私保护要求，查询者通过概率推理攻击获取真实结果的置信度应满足数据隐私概率阈值，因此尺度参数 b 应满足如下方程

$$\text{conf}(x)=F(\alpha)-F(-\alpha)=\int_{-\alpha}^{\alpha} f(u)du \leq T_{\text{prob}} \quad (11)$$

4.3.1 尺度参数分析

由于置信度 $\text{conf}(x)$ 的数学表达式根据位置参数 μ 的不同而具有多种表达形式，本文对式(11)进行分类分项。

1) 当位置参数 $\mu \leq -\alpha \leq \alpha$ 时，式(11)可以改写为

$$\begin{aligned} \text{conf}(x) &= \frac{1}{2} \left(\exp\left(\frac{\mu+\alpha}{b}\right) - \exp\left(\frac{\mu-\alpha}{b}\right) \right) \leq T_{\text{prob}} \\ &\rightarrow \exp\left(\frac{\mu+\alpha}{b}\right) \leq 2T_{\text{prob}} + \exp\left(\frac{\mu-\alpha}{b}\right) \\ &\rightarrow \frac{\mu+\alpha}{b} \leq \frac{\mu-\alpha}{b} + \frac{2T_{\text{prob}}}{\exp\left(\frac{\mu-\alpha}{b}\right)} \\ &\rightarrow \frac{\exp\left(\frac{\mu-\alpha}{b}\right)}{b} \leq \frac{T_{\text{prob}}}{\alpha} \\ &\rightarrow \frac{\mu-\alpha}{b} - \ln b \leq \ln \frac{T_{\text{prob}}}{\alpha} \\ &\rightarrow b^2 - \left(1 - \ln \frac{T_{\text{prob}}}{\alpha}\right)b + \alpha - \mu \geq 0 \end{aligned} \quad (12)$$

由于尺度参数 $b > 0$ ，解上述不等式可得

$$b \geq \frac{1 - \ln \frac{T_{\text{prob}}}{\alpha} + \sqrt{\left(1 - \ln \frac{T_{\text{prob}}}{\alpha}\right)^2 - 4(\alpha - \mu)}}{2}.$$

2) 当位置参数 $-\alpha \leq \mu \leq \alpha$ 时，式(11)可以

改写为

$$\begin{aligned} \text{conf}(x) &= 1 - \frac{1}{2} \left(\exp\left(\frac{\mu-\alpha}{b}\right) + \exp\left(\frac{-\alpha-\mu}{b}\right) \right) \leq T_{\text{prob}} \\ &\rightarrow \exp\left(\frac{\mu-\alpha}{b}\right) + \exp\left(\frac{-\alpha-\mu}{b}\right) \geq 2(1-T_{\text{prob}}) \\ &\rightarrow \exp\left(\frac{2\mu}{b}\right) + 1 \geq 2(1-T_{\text{prob}}) \exp\left(\frac{\mu+\alpha}{b}\right) \\ &\rightarrow \exp\left(\frac{2\mu}{b}\right) \geq \frac{\mu+\alpha}{b} + \ln 2(1-T_{\text{prob}}) \\ &\rightarrow \frac{2\mu}{b} \geq \ln(\mu+\alpha) - \ln b + \frac{b \ln 2(1-T_{\text{prob}})}{\mu+\alpha} \\ &\rightarrow b(\ln b) - \frac{b^2 \ln 2(1-T_{\text{prob}})}{\mu+\alpha} - b \ln(\mu+\alpha) + 2\mu \geq 0 \\ &\rightarrow \left[1 - \frac{\ln 2(1-T_{\text{prob}})}{\mu+\alpha}\right] b^2 - [1 + \ln(\mu+\alpha)]b + 2\mu \geq 0 \end{aligned} \quad (13)$$

由于尺度参数 $b > 0$ ，解上述不等式可得

$$b \geq \frac{[1 + \ln(\mu+\alpha)] + \sqrt{[1 + \ln(\mu+\alpha)]^2 - 8\mu \left[1 - \frac{\ln 2(1-T_{\text{prob}})}{\mu+\alpha}\right]}}{2 \left[1 - \frac{\ln 2(1-T_{\text{prob}})}{\mu+\alpha}\right]}$$

3) 当位置参数 $-\alpha \leq \alpha \leq \mu$ 时，同 2) 中解决步骤，式(11)可以改写为

$$b^2 - \left(1 + \ln \frac{\alpha}{T_{\text{prob}}}\right)b + \alpha + \mu \geq 0 \quad (14)$$

由于尺度参数 $b > 0$ ，解上述不等式可得

$$b \geq \frac{1 + \ln \frac{\alpha}{T_{\text{prob}}} + \sqrt{\left(1 + \ln \frac{\alpha}{T_{\text{prob}}}\right)^2 - 4(\alpha + \mu)}}{2}.$$

4.3.2 隐私保护参数 ϵ 配置

4.3.1 节根据攻击者置信度和隐私保护概率阈值推导出差分隐私保护中 Laplace 噪声分布的尺度参数的取值范围，根据尺度参数 $b = \frac{\Delta f}{\epsilon}$ 可以

很容易地推导出隐私保护预算参数的配置方法。

1) 当位置参数 $\mu \leq -\alpha \leq \alpha$ 时, 隐私保护参数 ε 配置的数学表达式为

$$b = \frac{\Delta f}{\varepsilon} \geq \frac{1 - \ln \frac{T_{\text{prob}}}{\alpha} + \sqrt{\left(1 - \ln \frac{T_{\text{prob}}}{\alpha}\right)^2 - 4(\alpha - \mu)}}{2} \rightarrow \varepsilon \leq \frac{2\Delta f}{\left(1 - \ln \frac{T_{\text{prob}}}{\alpha} + \sqrt{\left(1 - \ln \frac{T_{\text{prob}}}{\alpha}\right)^2 - 4(\alpha - \mu)}\right)} \quad (15)$$

2) 当位置参数 $-\alpha \leq \mu \leq \alpha$ 时, 隐私保护参数 ε 配置的数学表达式为

$$b = \frac{\Delta f}{\varepsilon} \geq \frac{1 + \ln(\mu + \alpha) + \sqrt{[1 + \ln(\mu + \alpha)]^2 - 8\mu \left[1 - \frac{\ln 2(1 - T_{\text{prob}})}{\mu + \alpha}\right]}}{2 \left[1 - \frac{\ln 2(1 - T_{\text{prob}})}{\mu + \alpha}\right]} \rightarrow \varepsilon \leq \frac{2\Delta f \left[1 - \frac{\ln 2(1 - T_{\text{prob}})}{\mu + \alpha}\right]}{\left(1 + \ln(\mu + \alpha) + \sqrt{[1 + \ln(\mu + \alpha)]^2 - 8\mu \left[1 - \frac{\ln 2(1 - T_{\text{prob}})}{\mu + \alpha}\right]}\right)} \quad (16)$$

3) 当位置参数 $-\alpha \leq \alpha \leq \mu$ 时, 隐私保护参数 ε 配置的数学表达式为

$$b = \frac{\Delta f}{\varepsilon} \geq \frac{1 + \ln \frac{\alpha}{T_{\text{prob}}} + \sqrt{\left(1 + \ln \frac{\alpha}{T_{\text{prob}}}\right)^2 - 4(\alpha + \mu)}}{2} \rightarrow \varepsilon \leq \frac{2\Delta f}{\left(1 + \ln \frac{\alpha}{T_{\text{prob}}} + \sqrt{\left(1 + \ln \frac{\alpha}{T_{\text{prob}}}\right)^2 - 4(\alpha + \mu)}\right)} \quad (17)$$

综上, 可以得到隐私保护预算参数 ε 配置的数学表达为

$$\varepsilon \leq \begin{cases} \frac{2\Delta f}{\left(1 - \ln \frac{T_{\text{prob}}}{\alpha} + \sqrt{\left(1 - \ln \frac{T_{\text{prob}}}{\alpha}\right)^2 - 4(\alpha - \mu)}\right)}, & \mu \leq -\alpha \leq \alpha \\ \frac{2\Delta f \left[1 - \frac{\ln 2(1 - T_{\text{prob}})}{\mu + \alpha}\right]}{\left(1 + \ln(\mu + \alpha) + \sqrt{[1 + \ln(\mu + \alpha)]^2 - 8\mu \left[1 - \frac{\ln 2(1 - T_{\text{prob}})}{\mu + \alpha}\right]}\right)}, & -\alpha \leq \mu \leq \alpha \\ \frac{2\Delta f}{\left(1 + \ln \frac{\alpha}{T_{\text{prob}}} + \sqrt{\left(1 + \ln \frac{\alpha}{T_{\text{prob}}}\right)^2 - 4(\alpha + \mu)}\right)}, & -\alpha \leq \alpha \leq \mu \end{cases} \quad (18)$$

从上述的分析可以推导出不同位置参数、尺度参数以及隐私概率阈值下隐私保护参数 ε 的选取范围。

4.3.3 不同查询权限下的隐私保护参数 ε 配置

本文提出由于查询者查询权限的不同, 在差分隐私保护中应根据查询者权限的不同对隐私

保护参数的配置进行调整。本文设定评估的查询权限取值范围为 $[0, 1]$, 为了向低查询权限用户配置较小的隐私保护预算参数, 设隐私保护预算参数 $\varepsilon' = \varepsilon Q_{\text{pri}}$, 基于此可以得到不同查询权限下隐私保护参数 ε' 的配置方法

$$\varepsilon' = \varepsilon_{Q_{\text{pri}}} \leq \begin{cases} \frac{2\Delta f Q_{\text{pri}}}{\left(1 - \ln \frac{T_{\text{prob}}}{\alpha} + \sqrt{\left(1 - \ln \frac{T_{\text{prob}}}{\alpha}\right)^2 - 4(\alpha - \mu)}\right)}, & \mu \leq -\alpha \leq \alpha \\ 2\Delta f \left[1 - \frac{\ln 2(1 - T_{\text{prob}})}{\mu + \alpha}\right] Q_{\text{pri}}, & -\alpha \leq \mu \leq \alpha \\ \frac{2\Delta f Q_{\text{pri}}}{\left(1 + \ln \frac{\alpha}{T_{\text{prob}}} + \sqrt{\left(1 + \ln \frac{\alpha}{T_{\text{prob}}}\right)^2 - 4(\alpha + \mu)}\right)}, & -\alpha \leq \alpha \leq \mu \end{cases} \quad (19)$$

通过这个隐私参数配置方法，可以设置隐私保护概率阈值并根据查询函数、查询者权限值，置信区间选取合适隐私保护参数 ε 达到数据隐私保护的目，保证数据效用最大化。所提出的隐私保护参数配置方法步骤如下。

步骤 1 提取查询用户的查询属性。对于每一个提交查询的查询用户计算其查询等级以及其查询函数的敏感度。

步骤 2 根据被查询数据的隐私属性设置隐私保护概率阈值 T_{prob} 。在数据集中不同隐私属性的数据所需要的隐私保护度是不同的，高敏感的隐私数据其隐私保护概率阈值较大；反之，隐私保护概率阈值较小。

步骤 3 根据噪声分布位置参数 μ 与置信区间 $[-\alpha, \alpha]$ 之间的对应关系计算攻击者概率推理攻击模型下的置信度 $\text{conf}(x)$ 。在对查询用户返回查询结果时，攻击者容易通过概率推理得出真实结果。为了更好地保护数据集中的用户敏感数据信息，本文对攻击者获取真实结果的把握，也就是置信度进行分析，通过设置隐私保护概率阈值、分析不同用户查询权限配置合适的隐私保护参数保护数据隐私信息。

步骤 4 计算隐私保护参数 ε 的最大值。对

于某个查询，通过 $\text{conf}(x) = F(\alpha) - F(-\alpha) =$

$\int_{-\alpha}^{\alpha} f(u) du \leq T_{\text{prob}}$ 以及式(19)计算隐私保护参数 ε

的配置范围。

4.4 数学实验分析

为了更加直观和清楚的表述问题，在本节中以计数查询为例，根据 4.3 节中的隐私保护参数的配置方法，对参数 ε 进行数学定性与定量分析。

如图 3 所示，对差分隐私保护机制中影响噪声分布的位置参数 μ 和攻击者置信区间 $[-\alpha, \alpha]$ 进行分类分析。当位置参数 μ 位于置信区间外时 ($\mu \leq -\alpha$ 或者 $\alpha \leq \mu$)，从图中可以看出在此情况下攻击者基于置信区间 $[-\alpha, \alpha]$ 的置信度较低，意味着攻击者不能有效地获取数据集中的真实信息，这是由于位置参数较大引起数据失真严重，数据的可用性较低。当位置参数 μ 位于置信区间内时 ($-\alpha \leq \mu \leq \alpha$)，攻击者置信度较高，对隐私保护的隐私分析具有可参考意义。

基于图 3 的分析，本文对位置参数 μ 位于置信区间内时不同置信区间长度对攻击者置信度的影响进行分析，同时根据上文所述的隐私保护参数配置方法设置不侵犯隐私的噪声参数保护水平。根据上文的数学分析以及图 4 所示，可以看

出隐私保护参数 ε 的配置与噪声分布的位置参数 μ 、置信区间 $[-\alpha, \alpha]$ 以及根据数据隐私保护属性所设置的隐私概率阈值有关。相同置信区间下, 当位置参数 μ 取 0 时攻击者置信度较大。在相同位置参数下, 攻击者置信区间越大, 置信度越大, 可以在不违背隐私保护概率阈值 T_{prob} 的前提下获取最大隐私保护参数值。

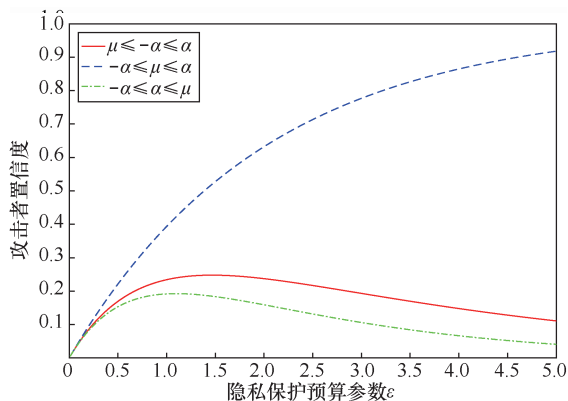


图 3 不同位置参数下的攻击者置信度

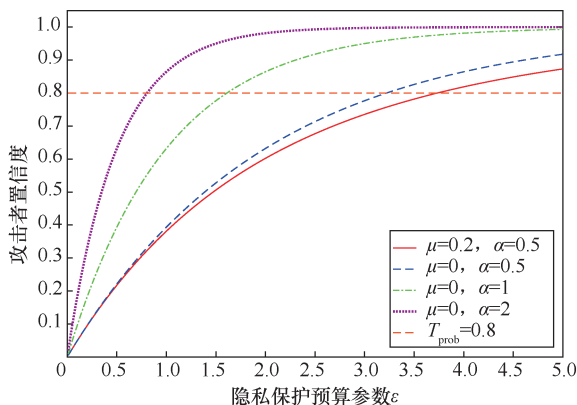


图 4 不同置信区间下的攻击者置信度

本文提出对于不同的查询用户其可能具有不同的查询权限, 即如果对所有查询用户提供相同隐私保护参数下的数据同样会造成隐私泄露的危险, 因此在 4.3 节隐私保护参数配置中增加了对查询用户查询权限的考虑。对于查询权限较高的查询者配置相对较小的隐私保护参数从而保证其获取较高可用性的数据。如图 5 所示, 对不同查询权限下的置信度进行分析, 从图中可以看出对于不同查询权限的查询用户, 在相同隐私保护概

率阈值下, 查询权限越小, 所配置的隐私保护预算参数 ε' 上限越小, 通过对隐私保护预算参数上限的限制可以实现对不同查询权限的查询用户配置不同取值范围的隐私保护参数, 达到细粒度的隐私保护效果。

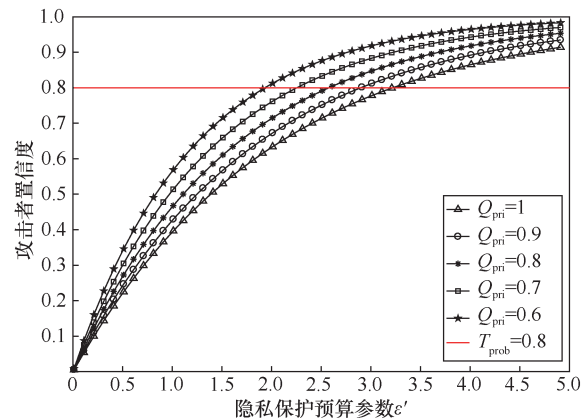


图 5 不同查询权限下的攻击者置信度

5 结束语

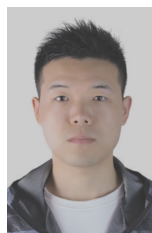
虽然差分隐私受到了广泛关注, 但如何配置合适的隐私保护参数仍然是个问题。本文证明了攻击者基于置信区间分析的概率推理攻击使攻击者能够从查询返回结果中通过使用噪声分布规律结合其置信区间设定推断查询的真实结果。为了解决上述问题, 本文通过对不同置信区间、噪声分布的位置参数以及尺度参数下的攻击者置信度进行分析, 设置隐私保护概率阈值以实现隐私披露概率的控制, 设计了一种基于置信度分析的差分隐私保护参数配置方法。从数学实验分析来看, 所提出的参数配置方法能够根据攻击者的置信度灵活地选取合适的隐私保护参数, 保证不同查询权限的查询者获取不同准确性的结果, 很好地解决了查询权限差异性带来的隐形披露问题。

参考文献:

- [1] LIANG F, YU W, AN D, et al. A survey on big data market: pricing, trading and protection[J]. IEEE Access, 2018, 6:15132-15154.

- [2] INAN A, GURSOY M E, SAYGIN Y. Sensitivity analysis for non-interactive differential privacy: bounds and efficient algorithms[J]. IEEE Transactions on Dependable & Secure Computing, 2017, (99):1-1.
- [3] LATANYA S. k-anonymity: a model for protecting privacy[J]. International Journal on Uncertainty, Fuzziness and Knowledge based Systems. 2002,10(5):557-570.
- [4] ASHWIN M, JOHANNES G, DANIEL K. ℓ -diversity: privacy beyond k-anonymity[C]. International Conference on Data Engineering. 2006,1(1): 24-35.
- [5] LI N H, LI T C, SURESH V. T-closeness: privacy beyond k-anonymity and l-diversity[C]. IEEE International Conference on Data Engineering. 2007:106-115.
- [6] DWORK C, ROTH A. The algorithmic foundations of differential privacy [M]. Now Publishers Inc. 2014.
- [7] ZHU T Q, LI G, ZHOU W, et al. Differentially private data publishing and analysis: a survey [J]. IEEE Transactions on Knowledge & Data Engineering, 2017, 29(8):1619-1638.
- [8] DWORK C. Calibrating noise to sensitivity in private data analysis [J]. Lecture Notes in Computer Science, 2012, 3876(8):265-284.
- [9] 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):101-122.
- XIONG P, ZHU T Q, WANG X F. A survey on differential privacy protection and application[J] Chinese Journal of Computers, 2014, 37(1):101-122.
- [10] LI Q, LI Y, ZENG G, et al. Differential privacy data publishing method based on cell merging[C]// IEEE International Conference on Networking, Sensing and Control. 2017:778-782.
- [11] SORUIA-COMAS J, DOMINGO-FERRER J, SANCHEZ D, et al. Individual differential privacy: a utility-preserving formulation of differential privacy guarantees [J]. IEEE Transactions on Information Forensics & Security, 2017, 12(6):1418-1429.
- [12] SMITH M, LOPEZ MAA, ZWIESSELE M, et al. Differentially private regression with Gaussian processes[C]// International Conference on Artificial Intelligence and Statistics. 2018.
- [13] HSU J, GABOARDI M, HAEBERLEN A, et al. Differential privacy: an economic method for choosing epsilon[C]//2014 IEEE 27th Computer Security Foundations Symposium. 2014: 398-410.
- [14] LEE J, CLIFTON C. How much is enough? choosing ϵ for differential privacy[C]//International Conference on Information Security. 2011: 325-340.
- [15] 何贤芒, 王晓阳, 陈华辉. 差分隐私保护参数 ϵ 的选取研究[J]. 通信学报, 2015, 36(12):124-130.
- HE X M, WANG X Y, CHEN H H. Study on choosing the parameter ϵ in differential privacy[J]. Journal on Communications, 2015, 36(12):124-130.
- [16] 郝晨艳, 彭长根, 张盼盼. 重复攻击下差分隐私保护参数 ϵ 的选取方法[J]. 计算机工程, 2018, 44(7): 151-155.
- HAO Y C, PENG C G, ZHANG P P. Selection method of differential privacy protection parameter ϵ under repeated attack[J]. Computer Engineering, 2018, 44(7): 151-155.
- [17] DWORK C. Differential privacy[J]. Encyclopedia of Cryptography and Security, 2011: 338-340.
- [18] DWORK C, ROTH A. The algorithmic foundations of differential privacy[J]. Foundations and Trends in Theoretical Computer Science, 2014, 9(3-4): 211-407.
- [19] BONOMI L, XIONG L. A two-phase algorithm for mining sequential patterns with differential privacy[C]// The 22nd ACM International Conference on Information & Knowledge Management. 2013: 269-278.

[作者简介]



李森有（1993—），河南驻马店人，国家数字交换系统工程技术研究中心硕士生，主要研究方向为新一代移动通信技术、移动通信安全、隐私保护。

季新生（1968—），男，河南驻马店人，国家数字交换系统工程技术研究中心教授、博士生导师，主要研究方向为网络空间安全、移动通信网络安全、拟态安全。

游伟（1984—），男，山东青岛人，国家数字交换系统工程技术研究中心讲师，主要研究方向为新一代移动通信网络技术、移动通信网络安全。