

差分隐私保护下的数据分级发布机制

张文静, 李晖

(西安电子科技大学 综合业务网理论与关键技术国家重点实验室, 陕西 西安 710071)

摘要: 隐私保护技术解决了数据发布过程中的隐私泄露问题, 然而当前的数据发布技术大多只面向查询用户发布同一隐私保护级别的数据, 并未考虑查询用户等级不同的情况。在所提出的满足差分隐私的数据分级发布机制中, 数据发布方利用隐私预算参数不同的拉普拉斯机制对数据查询结果进行隐私保护处理, 实现了输出隐私保护程度不同的查询结果。在依据付费或权限对查询用户分级后, 数据发布方为等级较高(低)的查询用户发布隐私保护程度较低(高)的查询结果, 使得查询用户可使用错误率较低(高)的数据, 达到了隐私数据分级发布的效果。实验结果与安全性分析表明该机制在抵抗背景知识攻击的同时还可有效地实现输出错误率不同的分级查询结果。

关键词: 差分隐私; 数据分级发布; 拉普拉斯机制; 隐私保护

中图分类号: TP309

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2015.00008

A differentially-private mechanism for multi-level data publishing

ZHANG Wen-jing, LI Hui

(State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China)

Abstract: Privacy preserving technology had addressed the problem of privacy leakage during data publishing process, however, current data publishing technologies mostly focused on publishing privacy preserving data with single level, without considering some scenarios of multi-level users. Therefore, a differentially-private mechanism for multi-level data publishing was proposed. The proposed mechanism employed the Laplace mechanism with different privacy budgets to output results with different privacy protection levels. After the user's level was determined according to the charge or privilege of that specific user, the goal that a user with high(low) level can only use the output result with low(high) privacy protection level which had low(high) error rate could be achieved. Finally, the evaluation results and security analysis show that our proposed framework can not only prevent from background knowledge attack, but also achieve multi-level data publishing with different error rates effectively.

Key words: differential privacy; multi-level data publishing; Laplace mechanism; privacy preserving

1 引言

随着信息技术的飞速发展以及大数据时代的到来, 数据的收集、发布与分析需求日益增多,

这些数据中潜在的用户隐私泄露问题也成为了人们广泛关注的焦点。当前, 大量的组织机构存储着不计其数的用户信息, 如表1所示, 这些数据除了可被合法拥有者使用外, 还会面向一些具有

收稿日期: 2015-09-07; 修回日期: 2015-09-27。通信作者: 张文静, xd.zhangwenjing@gmail.com

基金项目: 国家自然科学基金—广东联合基金资助项目(U1401251)

Foundation Item: The National Natural Science Foundation of China—Guangdong Provincial People's Government of the Joint Natural Science Fund Projects(U1401251)

表1 各组织机构产生含隐私信息的数据集

数据分类	数据拥有者	数据使用者	隐私信息	发布用途
医疗数据	医院	疾病研究学家	身份、疾病等	疾病成因研究
人口普查记录	政府	相关机构	年龄、收入等	生活状况调查
学籍信息	学校	教师	个人成绩等	学生情况调研
社交网络信息	企业	企业	购物记录等	市场前景预测

相应权限的机构发布以用于科学研究、人口调查、市场预测、广告推送等，这对政府管理、社会进步、企业发展等具有重要意义。然而，这些数据中可能包含极其敏感的私人信息，如疾病、收入、身份、兴趣及位置信息等，这些敏感信息的泄露将给用户自身带来不可估量的威胁和损失。因此，如何在数据发布时对用户敏感信息进行隐私保护成为了国内外相关领域的研究热点。

不同机构对于发布的数据具有不同的使用需求，如医疗数据中的疾病信息，负责病房分配的医生仅需了解病人疾病的大概信息即可分配病人入住到合适的病房，而主治医师则必须清楚病人确切的疾病信息以便于进行针对性治疗，这种场景下就需要对数据进行分级发布。实际中还存在很多需要发布不同隐私保护程度数据的场景，特别是当用户群中存在不同信任等级、不同访问权限、数据可用性需求不同的用户时。例如，政府机构相对于普通的未知机构具有更高的权限，因此，发布给政府机构一个隐私保护程度较低的数据版本是合理的，对于未知机构，则应当发布隐私保护程度较高的数据版本。此外，在基于数据隐私保护程度强弱进行收费的场景下，实行数据分级发布显得更加必要。

对于发布数据的使用，用户可能不希望泄露自己查询发布数据这一行为，因此，也有必要保护用户身份隐私。可见，研究如何在不泄露数据隐私的前提下，针对用户不同的访问权限或数据内容的不同需求，对用户进行等级划分进而实现数据分级发布，同时保护用户身份隐私，对在大数据时代实现安全、有效、合理的数据分析与使用具有重要意义。

隐私保护技术作为保护数据敏感信息、防止隐私泄露的安全技术，在学术研究领域得到了广

泛关注，该技术既保护了用户隐私，又提供了数据可用性。差分隐私保护算法(DP, differential privacy)^[1]作为一种新型、轻量级的隐私保护算法，具有严格的隐私定义且隐私保护粒度可控，近年来被应用到多种需要隐私保护的场景中。

考虑到数据发布过程中对数据隐私保护、查询用户身份隐私保护以及数据分级发布的需求，提出了满足差分隐私的数据分级发布机制，利用差分隐私算法的拉普拉斯机制对数据进行隐私保护处理。针对不同的用户需求或权限对用户分级，利用假名机制保护用户身份隐私，并通过在差分隐私算法中设置不同的隐私预算参数实现数据的分级发布。实验证明提出的方案具有一定的可行性。

2 相关工作

通常使用的隐私保护技术是将用户数据匿名化，使得用户隐私无法被识别出，如 k -anonymity^[2]，但是攻击者可使用背景信息对这种方法进行去匿名(de-anonymized)^[3,4]，从而分析出用户隐私信息。针对这种攻击模型，又相继提出了几种改善方法，如 l -diversity^[5]、 t -closeness^[6]等。尽管上述的隐私保护方法能够在一定程度上保护数据内容，但其缺陷在于对攻击者的背景知识和攻击模型做出了过多假设，不断出现的针对性攻击对这几种方法提出了巨大挑战。此外，这些隐私保护技术无法衡量数据内容的隐私保护程度，也不提供针对不同用户等级发布不同级别数据的功能。

由微软研究院的 Dwork 等人提出的差分隐私算法^[1,7]解决了上述难题。差分隐私算法主要通过查询的输出结果上添加噪声来保护数据隐私，使得查询者无法推测出某个单独用户或某组用户是否存在于数据库中，且不关心攻击者拥有背景知识的数量，定义了严格的攻击模型，克服了传

统隐私保护方法的缺陷。此外,差分隐私算法对于隐私保护水平也给出了严格的定义和量化评估方法,鉴于差分隐私算法的这种特性,当前已有许多针对差分隐私算法的应用^[8-11]。目前,已有多种基于差分隐私的数据发布算法,如直方图发布^[12,13]、划分发布^[14]、列联表发布^[15]等,基于差分隐私的分布式隐私数据查询也有一些相关研究^[16]。此外,也有针对特定场景中如何折中数据可用性与隐私保护程度^[17],以及提高差分隐私算法中数据可用性的研究^[18]。在数据发布时,经常使用两种实现机制来设计满足差分隐私定义的隐私保护算法,分别是拉普拉斯机制(Laplace mechanism)^[7]和指数机制(exponential mechanism)^[19]。二者主要的区别在于拉普拉斯机制适用于查询输出结果类型为数值型的隐私保护算法,而指数机制则主要适用于查询输出结果类型为非数值型的隐私保护算法。考虑到现有发布系统多数为直接发布,大多未涉及数据分级发布的情况,针对应用背景需求,结合差分隐私算法中隐私预算参数可控的特点,设计了基于差分隐私的数据分级发布机制,该机制可有效地实现对数据提供者的数据内容隐私保护,并按照用户的不同等级实现面向分级的发布机制。此外,考虑到保护数据使用者身份隐私的需求,引入了可信代理来实现数据使用者身份隐私保护。

3 差分隐私的理论基础

3.1 基本定义

差分隐私保护技术^[1,19]的基本思想是通过在数据查询的统计结果上添加噪声来实现隐私保护。这种方法确保了在添加或者删除某一项数据纪录时不会对查询的输出结果产生影响,使得经差分隐私算法处理操作后的数据集仍具有某些统计特性,即数据具有一定的可用性。差分隐私的定义描述如下。

定义1 差分隐私^[2]。

给定邻近数据集 D_1 与 D_2 ,即二者之间至多相差一条记录,设有随机算法 A , P_A 为包含算法 A 所有可能输出结果的集合。对于任意两个给定的邻近数据集 D_1 与 D_2 以及 P_A 的任何子集 S_A ,若算法 A 满足式(1),则称算法 A 提供 ϵ -差分隐

私保护。

$$P_r[A(D_1) \in S_A] \leq e^\epsilon \times P_r[A(D_2) \in S_A] \quad (1)$$

其中,概率 P_A 由算法 A 决定,表示隐私的泄露风险; ϵ 为隐私保护预算,表示算法 A 可提供的隐私保护程度,通过设置参数 ϵ 来保证在数据集中删除任何一条数据时,算法输出同一结果的概率不发生明显变化。 ϵ 取值越小,隐私保护程度就越高。在算法的实际使用当中,时常将 ϵ 取为较小的值,如0.01、0.1、ln2、ln3等。鉴于隐私保护预算 ϵ 的这种特性,可以通过设置 ϵ 的不同取值来实现算法输出结果隐私性与数据可用性的折中。

3.2 拉普拉斯机制

拉普拉斯机制通过向查询的真实输出值中加入由拉普拉斯分布产生的随机噪声来实现 ϵ -差分隐私保护,该机制主要适用于查询输出结果类型为数值型的隐私保护算法。

定义2 拉普拉斯机制^[11]。

给定已知数据集 D ,设有查询函数 $f: D \rightarrow R^d$,其敏感度为 Δf ,则有隐私保护算法 $A(D) = f(D) + N$ 提供 ϵ -差分隐私保护。其中, $N \sim Lap(\Delta f/\epsilon)$ 表示 N 是服从尺度参数(scale parameter)为 $\Delta f/\epsilon$ 的拉普拉斯噪声分布;函数敏感度 Δf 是指删除数据集当中任一条记录后,该删除操作对输出结果产生的最大影响。

在加入到查询输出结果的拉普拉斯分布中,当位置参数为0时,不同的尺度参数对应着不同的拉普拉斯噪声分布,进而对查询输出结果产生不同影响。设位置参数为0、尺度参数为 b 的拉普拉斯分布 $Lap(b)$ 的概率密度函数为 $p(x)$,其表示形式如下

$$p(x) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right) \quad (2)$$

图1为尺度参数 b 分别取1、2、4、8时对应的拉普拉斯分布曲线。由于差分隐私的拉普拉斯机制添加到输出结果中的噪声为形如 $N \sim Lap(\Delta f/\epsilon)$ 的分布函数,即尺度参数 $b = \Delta f/\epsilon$ 。可见,当函数敏感度 Δf 固定时, ϵ 与 b 成反比,即 ϵ 越小,尺度参数 b 越大,曲线越趋向于扁平,添加到输出结果中的噪声越大,达到的隐私保护程度就越高,但数据可用性会随之降低。因此,可

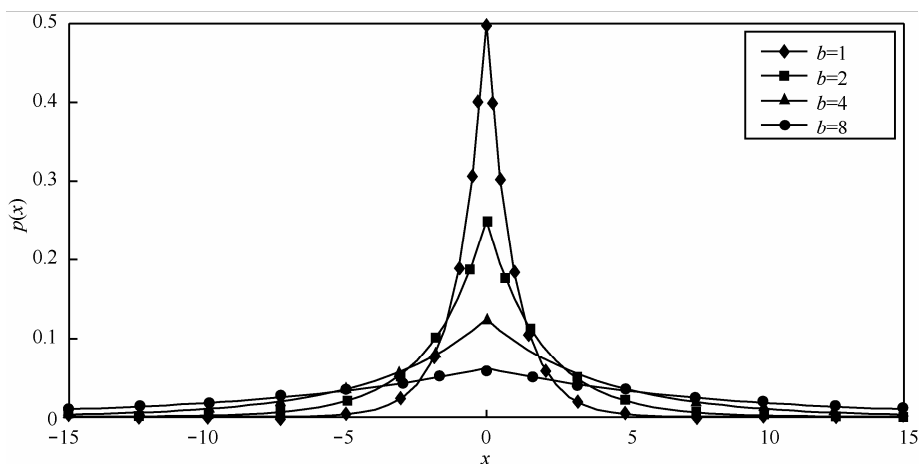


图1 不同参数的拉普拉斯概率密度函数

依据拉普拉斯机制的这种特性来实现隐私数据的分级发布。

4 基于差分隐私的数据分级发布机制

4.1 基本框架

为实现隐私保护的数据发布，提出的隐私保护框架采用差分隐私算法对发布前的数据进行隐私保护处理。此外，根据查询用户的付费或权限情况进行用户等级划分，通过合理设置差分隐私算法中的隐私预算参数 ϵ 来实现隐私数据的分级发布。满足差分隐私的数据分级发布框架由3部分构成，分别是：查询用户、可信代理以及隐私数据分级发布系统，如图2所示。

为了使查询用户在使用数据时不泄露身份，在该框架中加入了使用假名机制(pseudonymity)的可信代理来实现用户身份隐私保护，假名机制确保了用户在不泄露身份的前提下使用某种资源或服务。在系统初始化阶段，每个需要使用查询服务的用户在可信代理处注册身份，可信代理利用假名机制为该用户分配一个与其身份对应的假名标识符，并依据用户需求、付费情况或权限进行等级划分（付费多或权限高则分配高等级），通过关联用户身份假名与等级生成“身份假名—等级”数据集，存储格式如表2所示。其中，用户身份标识记为 ID ，身份假名标识记为 $Pseu(ID)$ ，付费或权限记为

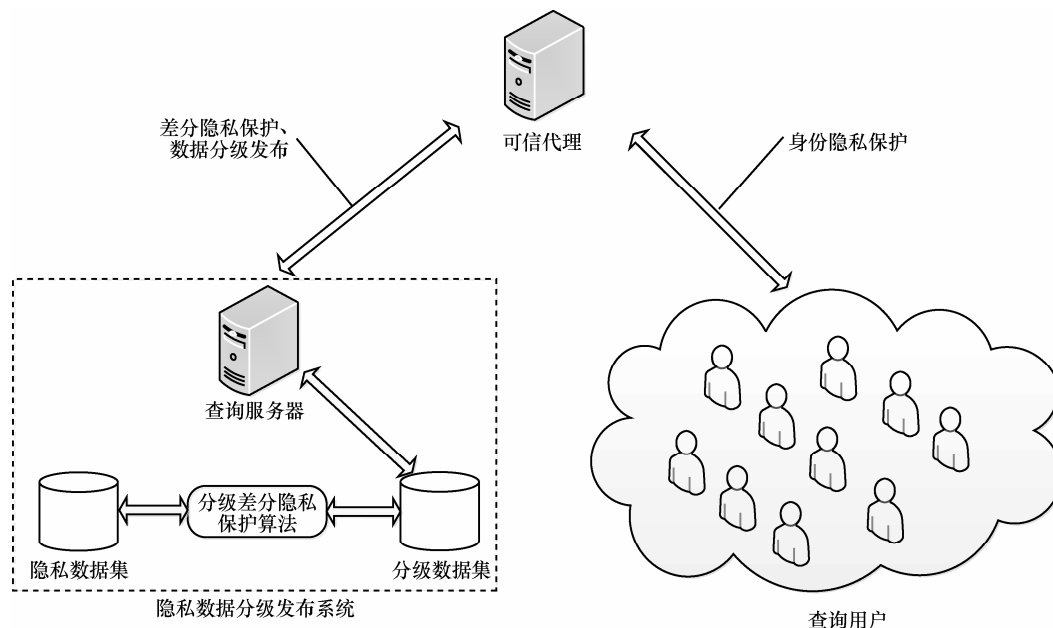


图2 差分隐私保护下的数据分级发布系统框架

P , 相应等级记为 L , ε 为差分隐私算法中的隐私预算参数。

表2 “身份假名—等级”数据集存储格式

身份假名标识符	付费或权限	查询用户等级	ε
$Pseu(ID_1)$	P_1	L_1	ε_1
$Pseu(ID_2)$	P_2	L_2	ε_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$Pseu(ID_n)$	P_n	L_n	ε_n

可信代理将该数据集发送至查询服务器存储。数据发布系统根据用户身份的对应等级,向其发布具有相应隐私保护程度的统计数据或查询结果。用户在后续的查询操作中,只需向可信代理发送身份 ID ,由可信代理将身份转换为假名后再发送至查询服务器。查询服务器收到该假名时,会在“身份假名—等级”数据集中进行搜索匹配,若匹配成功,则返回由与其等级对应的 ε -差分隐私算法处理后的隐私数据(查询结果);若匹配失败,则返回查询失败的通知消息。可信代理收到消息后,在本地查找与假名对应的真实用户身份,再将查询结果转发至对应用户。查询操作流程示意如图3所示。

4.2 隐私数据分级发布算法

在隐私数据分级发布系统中,利用差分隐私算法的拉普拉斯机制实现数据隐私保护,向查询结果中加入符合拉普拉斯分布的噪声,即查询输出为 $A(D) = f(D) + Lap(\Delta f/\varepsilon)$, $f(D)$ 为查询函数的真实输出。依据用户不同等级,选择不同隐私预算参数 ε ($\varepsilon > 0$) 对应的隐私保护算法,实现输出结果分级化。隐私数据分级发布算法如图4所示。

该算法实现了 ε -差分隐私。

- 1) 根据用户 i 付费情况或权限确定用户等级 L_i ;
- 2) 选择与 L_i 对应的 ε , 生成符合 $Lap(\Delta f/\varepsilon)$ 分布的噪声 n_i ;
- 3) 将噪声加入查询结果得到 $A_i(D) = f_i(D) + Lap(\Delta f/\varepsilon)$;
- 4) 返回加噪后的查询结果 $A_i(D)$ 至用户 i 。

图4 隐私数据分级发布算法

证明 考虑到所有的数据集 D_1 和 D_2 至多相差一条记录,且向真实查询结果中添加的噪声符合拉普拉斯分布,因此由式(2)有 $\frac{P_r[A(D_1) \in S_A]}{P_r[A(D_2) \in S_A]} = e^{\frac{-|f(D_1) - f(D_2)|\varepsilon}{\Delta f}}$ 。由于查询操作的函数敏感度 $|f(D_1) - f(D_2)| \leq \Delta f$, 则有 $\frac{P_r[A(D_1) \in S_A]}{P_r[A(D_2) \in S_A]} \leq e^\varepsilon$, 由定义1可知 $A(D) = f(D) + Lap(\Delta f/\varepsilon)$ 符合 ε -差分隐私。

数据分级发布机制在差分隐私算法中设置不同的隐私预算参数 ε , 实现了数据隐私保护程度的可控性以及查询结果与用户级别的对应性,因此在分级数据集中存储了不同 ε 对应的差分隐私算法处理后的隐私数据。用户等级越高,拉普拉斯机制中使用的隐私预算参数 ε 越大,数据隐私保护程度越低,查询结果错误率越低,数据可用性也就越高,反之亦然。

可见,用户等级间的本质区别在于返回其查询结果的准确度不同,为保证用户等级划分的合理性,可依据授权或付费等方式来划分用户等级。例如,在付费场景中,可按用户付费由多到少为其分配由高至低的等级,获得的查询结果的准确度也由高到低;在授权场景中,可将用户分为直接授权、间接授权以及未授权3种等级,直接授权的用户访问权限最高,获得的查询结果的准确度最高,未授权的用户则只能获得准确度最低的查询结果。例如,若 ε 从集合 $(0.01, 0.4)$ 中取值,



图3 查询操作流程

则可为级别最高（付费最多/权限最高）的用户发布 $\epsilon = 0.4$ 的拉普拉斯机制处理的结果，而为级别最低（付费最少/权限最低）的用户发布 $\epsilon = 0.01$ 的拉普拉斯机制处理的结果。

5 实验结果与安全性分析

5.1 实验结果

实验中，将用户的查询操作设置为计数查询（计数查询的函数敏感度为 1），利用差分隐私的拉普拉斯机制实现查询结果的隐私保护，隐私预算参数 ϵ 的值取自集合 $(0.001, 0.1)$ 。设查询结果的真实输出值为 V_{true} ，隐私保护处理后的查询结果为 V_{noise} ，定义错误率为 $\frac{|V_{\text{true}} - V_{\text{noise}}|}{V_{\text{true}}}$ 。实验中对于每个特定 ϵ 值对应的差分隐私算法进行 20 次计数查询，再对这些查询结果的错误率取平均值，该值与 ϵ 的对应关系如图 5 所示。

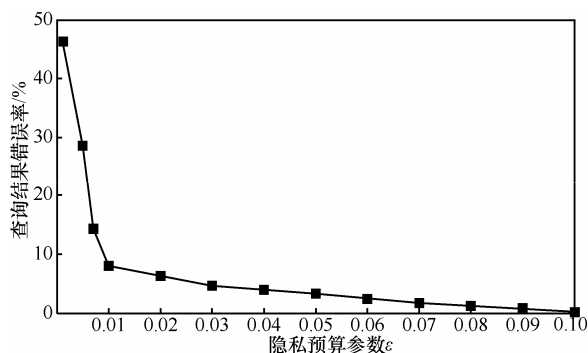


图5 隐私预算参数 ϵ 与查询结果错误率的对应关系

由图 5 可以看出，随着 ϵ 的增大，数据隐私保护程度的效果减弱，使得查询结果的错误率降低，数据准确率或数据可用性则越高。因此，可将查询结果错误率作为标准来对数据发布等级进行划分，等级最高的用户可访问错误率小于 1% 的数据（可由 ϵ 值为 0.1 对应的差分隐私算法生成），而等级最低的用户则只能访问错误率为 20%~30% 之间的数据（可由 ϵ 值为 0.005 对应的差分隐私算法生成），进而实现了数据发布等级的有效划分。

对于 ϵ 的分级标准，可依据发布数据错误率来衡量。若希望发布错误率小于 1% 的数据，则选择 $\epsilon = 0.1$ 的拉普拉斯机制；若希望发布错误率在 20%~30% 之间的数据，则选择 $\epsilon = 0.005$ 的拉普拉

斯机制。表 3 是一种可能的依据 ϵ 对应的查询结果错误率进行用户等级划分的方式（ L_1 为最低等级， L_7 为最高等级）。

表 3 一种可能的用户等级划分方式

ϵ	错误率	用户等级
0.001	46.5%	L_1
0.005	28.6%	L_2
0.007	14.3%	L_3
0.01	8.0%	L_4
0.03	4.7%	L_5
0.07	1.8%	L_6
0.1	0.3%	L_7

在一些场景中，依据查询结果错误率来划分用户等级是合理的。例如，假设存在多个实体需要通过查询人口普查记录获得某城市 25 岁至 35 岁之间人口的平均收入。政府部门为了宏观调控或市场定位等目的需要较为精确的查询结果，其通常具有人口普查记录的最高访问权限，因此可查询到错误率最低（错误率 < 1%）的收入值；普通人员则权限较低，只能查询出大概的收入值（错误率 $\approx 15\%$ ），但这一查询结果对于普通人员来说是足够的，因为他可能只需要了解收入的大概范围用以判断该城市的生活标准，无需获得精确的收入值。

5.2 安全性分析

1) 分级的差分隐私保护

通过差分隐私的拉普拉斯机制实现了数据隐私保护，并通过将用户等级与隐私预算参数 ϵ 关联来实现数据的分级发布，用户获得查询结果的错误率与其等级直接相关。实现了隐私保护程度分级、查询输出结果可控的数据发布机制。

2) 背景知识攻击

由于差分隐私在定义算法时进行了最大背景知识假设，其假设攻击者即使已掌握除某条数据记录之外的所有记录信息，也能保证该条数据记录的隐私不被泄露。因此，差分隐私保护下的数据发布机制可有效地抵抗背景知识攻击。

3) 身份隐私保护

由于在本文的隐私保护机制中引入了可信代理，使得用户不再与服务器直接通信，可信代理

利用假名将用户真实身份替换后再发送至服务器,保证了用户在服务器端不会泄露身份,进而保护了用户的身份隐私。

4) 函数敏感度对隐私保护效果的影响

拉普拉斯机制通过在输出结果中添加服从 $Lap(\Delta f/\epsilon)$ 分布的噪声实现差分隐私,噪声大小由函数敏感度 Δf 和隐私预算参数 ϵ 共同确定,即除 ϵ 外,函数敏感度也会影响隐私保护效果。对于相同的 ϵ , 函数敏感度 Δf 越大,加入噪声越多,隐私保护程度也就越高,查询结果错误率越高。在实验中,使用计数查询作为函数 f ,由于计数查询的函数敏感度为 1,即在拉普拉斯机制中加入的噪声方差为 $2/\epsilon^2$,此时隐私保护程度完全由隐私预算参数 ϵ 确定。当查询操作的函数敏感度较高时,为了达到相同的隐私保护效果,需引入更多的噪声,导致数据可用性降低。因此,在支持多种查询的数据发布系统中,也应考虑函数敏感度对隐私效果的影响。

6 结束语

大数据时代产生了大量有价值的信息,如何在有效使用这些数据的同时又实现数据隐私保护成为数据发布领域的研究热点。差分隐私作为一种严格的、可证明的隐私保护技术引起了人们的广泛关注,且通过设置隐私预算参数可使得发布数据具有不同的隐私保护程度。考虑到用户对发布数据使用需求的不同,迫切需要划分用户等级以实现发布数据的可用性与用户需求相对应。差分隐私算法的特性为隐私数据分级发布提供了理论基础,且拉普拉斯机制也为数据发布提供了支持差分隐私算法的实现机制。此外,用户通过可信代理进行查询操作,避免了用户身份在服务器端的泄露风险。本文旨在利用差分隐私技术实现数据的分级发布机制,建立数据分级发布框架,并对用户进行分级,仅为用户分配与等级对应的查询结果。该机制使得数据发布在保护隐私的同时,更加具有针对性。

在未来的工作中,将考虑在差分隐私保护下实现除计数查询外的其他查询结果的分级发布,并对算法进行优化,实现支持多种查询、高效、隐私保护的数据分级发布机制。在数据可用性方

面,将研究如何在提供同等隐私保护程度的同时降低查询结果错误率。此外,也将在用户等级划分部分中引入信任机制和访问控制机制,进一步细化、合理化等级划分方法。

参考文献:

- [1] DWORK C. Differential privacy[C]//Proceedings of the 33rd International Colloquium on Automata, Languages and Programming, Venice. 2006: 1-12.
- [2] SWEENEY L. K -anonymity: a model for protecting privacy[J]. International Journal on Uncertainty, Fuzziness and Knowledge Based Systems, 2002, 10(5): 557-570.
- [3] NARAYANAN A, SHMATIKOV V. Robust de-anonymization of large sparse datasets[C]//Proceedings of IEEE Symposium on Security and Privacy, Oakland. 2008: 111-125.
- [4] RIBEIRO B, CHEN W, MIKLAU G, et al. Analyzing privacy in enterprise packet trace anonymization[C]//Proceedings of the Network and Distributed System Security Symposium, San Diego. 2008.
- [5] MACHANAVAJJHALA A, KIFER D, GEHRKE J, et al. L -diversity: privacy beyond k -anonymity[C]//Proceedings of the 22nd International Conference on Data Engineering, Washington, DC. 2006: 24-35.
- [6] LI N, LI T, VENKATASUBRAMANIAN S. T -closeness: privacy beyond k -anonymity and L -diversity[C]//Proceedings of the 23rd International Conference on Data Engineering, Washington DC. 2007: 106-115.
- [7] 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, New York. 2006: 265-284.
- [8] ANDRES M E, BORDENABE N E, CHATZIKOKOLAKIS K, et al. Geo-indistinguishability: differential privacy for location-based systems[C]//Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security, Berlin. 2013: 901-914.
- [9] SHEN E, YU T. Mining frequent graph patterns with differential privacy[C]//Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago. 2013: 545-553.
- [10] ZHAO J, JUNG T, WANG Y, et al. Achieving differential privacy of data disclosure in the smart grid[C]//Proceedings of IEEE INFOCOM 2014, Toronto. 2014: 504-512.
- [11] XIAO Y, XIONG L. Protecting locations with differential privacy under temporal correlations[C]//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver. 2015: 1298-1309.
- [12] XU J, ZHANG Z, XIAO X, et al. Differentially private histogram publication[C]//Proceedings of IEEE 28th International Conference on Data Engineering, Washington DC. 2012: 32-43.
- [13] LI H, XIONG L, JIANG X, et al. Differentially private histogram publication for dynamic datasets: an adaptive sampling ap-

- proach[C]//Proceedings of the 24th ACM International Conference on Information and Knowledge Management, Melbourne. 2015: 1001-1010.
- [14] PENG S, YANG Y, ZHANG Z, et al. DP-tree: indexing multidimensional data under differential privacy[C]//Proceedings of the ACM SIGMOD International Conference on Management of Data, Scottsdale. 2012: 864-864.
- [15] YAROSLAVTSEV G, CORMODE G, PROCOPIUC C M, et al. Accurate and efficient private release of data cubes and contingency tables[C]//Proceedings of 29th IEEE International Conference on Data Engineering, Brisbane. 2013: 745-756.
- [16] CHEN R, REZNICHENKO A, FRANCIS P, et al. Towards statistical queries over distributed private user data[C]//Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation, San Jose. 2012: 169-182.
- [17] TRAMER F, HUANG Z, HUBAUX J P, et al. Differential privacy with bounded priors: reconciling utility and privacy in genome-wide association studies[C]//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver. 2015: 1286-1297.
- [18] SANCHEZ D, DOMINGO-FERRER J, MARTINEZ S. Improving the utility of differential privacy via univariate microaggregation[C]//Proceedings of Privacy in Statistical Databases, Ibiza. 2014: 130-142.
- [19] MCSHERRY F, TARWAR K. Mechanism design via differential privacy[C]//Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science, Providence, Rhode Island. 2007: 97-103.

作者简介:



张文静(1988-), 女, 黑龙江绥化人, 西安电子科技大学博士生, 主要研究方向为大数据安全、隐私保护。



李晖(1968-), 男, 河南灵宝人, 博士, 西安电子科技大学教授、博士生导师, 主要研究方向为密码学、无线网络安全、云计算安全、信息论与编码理论。