一种基于区块链的差分隐私成本节省和跟踪方法

杨钊，IEEE 研究生会员，IEEE 康会员，张泽航，，Dusit Niyato 电气和电子工程师学会会员史淑玉和林国彦 电气和电子工程师协会高级会员

*摘要—出于分析目的，现在收集的用户敏感信息越来越多。*为了保护用户信息的隐私，文献中对差别隐私进行了广泛的研究。隐私参数限制了噪声输出泄露的数据集信息。通常，一个数据集需要用于回答多个查询，因此随着更多的查询被回答，隐私保护级别可能会降低。因此，跟踪隐私预算支出至关重要，该支出不应超过隐私预算的给定限额。此外，如果一个查询之前已经被回答，并且在相同的数据集上被再次询问，我们可以将先前的有噪声的响应用于当前的查询，以节省隐私成本。鉴于以上情况，我们设计了一种算法，如果重复询问相同的查询，可以重用以前的有噪声的响应。特别地，考虑到同一查询的不同请求可能具有不同的隐私要求，我们的算法可以设置旧的有噪声响应的最优重用分数，并添加新的噪声以最小化累积的隐私成本。此外，我们设计并实现了一个基于区块链的系统，用于跟踪和节省差分隐私成本。因此，数据集的所有者将完全了解数据集是如何使用的，并确信一旦指定的隐私预算用尽，回答查询不会产生新的隐私成本。

索引术语—区块链、数据分析、差异隐私、高斯机制。

手稿于 2020 年 7 月 19 日收到；2020 年 10 月 21 日和 2020 年 12 月 22 日修订；2021 年 1 月 9 日受理.出版日期 2021 年 2 月 9 日；当前版本日期 2021 年 5 月 21 日。杨钊和赵骏的工作得到了南洋理工大学(NTU)启动基金的部分资助；部分由阿里巴巴-NTU 新加坡联合研究院(JRI)完成；部分由新加坡教育部学术研究基金第 1 级 RG128/18、第 1 级 RG115/19、第 1 级 RT07/19、第 1 级 RT01/19 和第 2 级 MOE2019-T2-1-176 资助；部分由 NTU-WASP 联合项目资助；部分由新加坡国家研究基金会(NRF)根据其战略能力研究中心资助计划资助:隐私保护技术和系统战略研究中心；部分由 NTU 能源研究所提供；部分由新加坡 NRF 国家卓越卫星、安全关键基础设施设计科学与技术国家统计局(NSoE DeST-SCI2019-0012)资助；部分由人工智能新加坡(AISG) 100 实验(100E)项目资助；部分由 NTU 大型垂直起降研究平台项目资助。Dusit Niyato 的工作得到了新加坡国家研究基金会(NRF)的部分支持，具体是通过新加坡能源市场管理局(EMA)、NRF2017EWT-EP003041 赠款下的能源弹性、NRF2015-NRF-ISF001-2277 赠款下的新加坡 NRF、NSoE DeST-SCI2019-0007 赠款下的新加坡 NRF 国家卓越卫星、安全关键基础设施设计科学和技术；部分由 A*STAR-NTU-SUTD 制造业未来人工智能联合研究资助(授予 RGANS1906)、瓦伦堡 AI、自主系统和软件计划和南洋理工大学(WASP/NTU)资助(授予 m 4082187(4080)；部分由新加坡教育部第一级(RG16/20)；部分由阿里巴巴集团通过阿里巴巴创新研究(AIR)计划和阿里巴巴-NTU 新加坡联合研究院(JRI)完成。林国彦的工作得到新加坡国家研究基金会战略能力研究中心资助计划的支持。*(对应作者:赵骏。*

杨钊、、康、张泽航、杜希特·尼雅托和林国彦在新加坡南洋理工大学计算机科学与工程学院(电子邮件:s 180049 @ e . NTU . edu . SG；赵骏@ NTU . edu . SG；kavinkang @ NTU . edu . SG；prestonzzh@163.com；dni yato @ NTU . edu . SG；kwokyan.lam@ntu.edu.sg)。史淑玉，南京大学计算机科学与工程系，中国南京 210008(电子邮件:ssy@nju.edu.cn)。

数字对象标识符 10.1109/JIOT . 2021 . 30582020209

爱达荷（Idaho 的缩写）。介绍

M正在收集大量用户敏感信息，用于数据分析和机器学习，如大规模物联网数据。一些物联网数据包含用户的机密信息，例如能耗或位置数据。它们可能会暴露一个家庭的习惯。为了保护个人隐私，许多国家对科技公司如何收集和处理用户数据有严格的政策。然而，公司需要分析用户的数据来提高服务质量。为了在揭示数据集有用信息的同时保护隐私，人们提出了差分隐私[6]–[8]。直观地说，通过加入一些噪声，DP 下的算法输出不会因为数据集中存在或不存在一个用户的信息而发生显著变化。由于其介绍[6]，[7]，动态规划引起了学术界[9]–[13]和工业界[14]–[16]的极大兴趣。比如苹果已经将 DP 纳入其移动操作系统 iOS[14]；谷歌已经在 Chrome 浏览器中实现了一个名为 RAPPOR 的 DP 工具来收集信息[15]。

粗略地说，实现(，δ)-DP [6]的随机化机制意味着，除了(通常很小的)概率 δ 外，更改数据库中的记录不能改变输出被看到的概率超过乘法因子 e。因此，关于由(，δ)-DP 算法的噪声输出泄露的数据集的信息受隐私参数和 δ 的限制。更小和 δ 意味着更强的隐私保护和更少的信息泄露。注意，非零信息泄漏是实现非零效用的必要条件。通常，数据集可以用于回答多个查询(例如，。对于多个分析任务)，从而积累了信息泄露，降低了隐私保护级别，这可以直观地理解为隐私支出的增加。因此，有必要记录隐私成本，以防止其超出隐私预算。当一个不同的私有方案被应用到现实世界的应用中时，隐私预算被用来量化隐私风险。此外，如果查询以前被回答过，我们通过重用旧的有噪声的响应来回答当前的查询来降低隐私成本。

传统上，回答对数据集的查询所产生的隐私成本由数据集持有者来承担。其信息在数据集中的用户不清楚使用情况。隐私消费有可能已经超出隐私预算。为了解决这个问题，新兴的区块链技术提供了一种新的解决方案来管理隐私成本。区块链是在不使用集中式服务器的情况下存储加密和防篡改交易记录的区块链[17]，[18]。随着区块链记录数据集如何用于回答查询，用户完全知道他们的信息是如何分析的。用户可以轻松访问区块链，查看隐私预算的消耗情况。数据集持有人有动机采用我们基于区块链的方法，为其信息在数据集内的用户提供以下责任保证:如果

数据集持有人使用的数据集多于区块链记录的查询集，则可以采取措施以发现数据集持有人作弊，因为写入区块链的交易是防篡改的。杨等。[19]提议利用区块链跟踪动力定位预算，但他们没有提出一个重复使用噪音的机制。相比之下，我们设计了一种动态规划机制来有效地重用先前查询的结果，从而重用噪声并降低隐私成本。在第六节，我们提出了更详细的比较。鉴于上述情况，我们提出了一种基于区块链的算法 1，并将其用于跟踪和管理差异化竞争成本，该算法使用区块链使隐私支出对数据所有者透明。因此，数据所有者可以通过检查区块链交易的信息来跟踪数据集的使用情况，包括每个查询的类型、用于回答每个查询的噪声响应、添加到真实查询结果的相关噪声水平以及剩余的隐私预算。除了提供隐私管理的透明度，我们的区块链系统的另一个优势如下。一旦指定的隐私预算用尽，在区块链实施的智能合同将确保不会产生新的隐私成本，这是可以验证的。此外，由于区块链存储了用于回答每个查询的噪声响应，因此我们还设计了一种算法，如果再次询问相同的查询，则通过重用先前的噪声响应来最小化累积的隐私成本。我们的算法(通过严格的证明)能够设置旧的有噪声响应的最佳重用分数，并添加新的噪声(如果必要的话)，考虑到同一查询的不同请求可能以不同的隐私要求发送。在我们基于区块链的系统中，重用有噪声的响应不仅节省了隐私成本，而且在生成有噪声的响应时无需联系托管数据集的服务器，从而减少了通信开销。

贡献:本文的主要贡献总结如下。

1)设计了一种新的隐私保护算法，该算法具有严格的数学证明，如果接收到相同的查询，则通过重用先前有噪声的响应，在有限的隐私预算下最小化累积的隐私成本。

表一
符号概述

因此，数据集可以用于回答更多的查询，同时防止隐私泄露，这对于频繁查询的数据集是必不可少的。病历数据集。

2) 我们设计的方法通过利用记录的噪声结果，大大减少了请求服务器的次数。

3) 我们根据详细的序列图实现了所提出的系统和算法，并使用真实数据集进行了实验。数值结果表明，我们提出的系统和算法在保持准确性的同时，有效地节省了隐私开销。

*组织:文章的其余部分组织如下。*第二节介绍了迪拜港口和区块链的初步情况。第三部分介绍了系统设计，包括我们提出的噪声重用算法。第四节描述了实施我们系统的挑战。在第五部分，我们讨论了实验结果，以验证我们的系统的有效性。第六节调查相关工作。第七节讨论了我们提出的计划的假设和限制。此外，我们确定了一些未来的方向。第八节总结了本文。

*符号:在本文中，P[ ]表示概率，F[ ]表示概率密度函数。*符号 N(0，A)表示均值和方差为零的高斯随机变量，当用于生成有噪声的查询响应时，表示新的高斯噪声。表一总结了本文其余部分使用的符号。

二. 预赛

我们将初赛的这一部分组织如下。在...里第二部分，我们介绍了动态规划的正式定义。在第二节中，我们解释了区块链、以太琴和智能契约的概念。

A. 差别隐私

DP 直观地意味着对手不能很有把握地确定随机输出是否到来

算法 1 我们提出的算法来回答第 m 次查询和调整剩余隐私成本

输入: √ d: 数据集；质量管理: 这 第 m 次 查询； (m，δm): 请求 隐私 参数√ 为 询问 QM；

（_ squared _ restrict _ budget，δbudget):剩余隐私预算(开头是(_平方_预算，δ 预算)

_平方_预算 = budget 2)；Qm: 2 查询 Qm 的灵敏度；

输出:Qm(D):在(m，δm)-DP 下对数据集 D 上的查询 Qm 的噪声查询响应；

—— —— —— —— 1: σm ←高斯(Qm，m，δm)；//注释:从引理 1 出发，认为高斯(Qm，m，δm)× Qmm。.

2:如果第一次看到查询 Qm，那么 √ =

3:客户端计算成本的平方， _平方_成本，δ 预算) σm；

4:$\sqrt{2\ln\frac{1.25}{\delta}}\times\frac{\Delta_{Qm}}{\epsilon_squared_cost}$评论:这意味着 δ 预算 √ $\sqrt{2\ln\frac{1.25}{\delta_m}}$ = σm，其中 σm 为高斯型(Qm，m，δm) × Qmm。

5:客户端计算_平方_剩余_预算↓_平方_剩余_预算_平方_成本；6:如果_平方_剩余_预算≥ 0，则

7:返回 Qm(D) ← Qm(D) + N(0，1)×σm；//注释:我们参考第三节-D 中的案例 1)。*如果 Qm 是多维的，独立的高斯噪声将被添加到每个维度。*

8:区块链记录 Qm 的查询类型，m，δm，σm，Qm(D)；//注释:这些信息将与数据集 D 的加密散列一起保存，Blockchain 存储这些散列，以便它知道哪些记录是针对同一数据集 D 的。

9: 其他

10: 返回隐私预算不足的错误；

11: 如果…就会结束

12:否则

13:假设 Qm 是一个 t 型查询。Blockchain 将 σm 与 t 中的值进行比较:= {σj:σj 已记录在 Blockchain 中，并且 Qj 是 t-query 类型}(即，。*t 由 t-query 类型的先前实例的相应噪声量组成)，导致以下子情况。*

14: 如果存在 σj ∈ t σm =σj 然后

15: 区块链返回 Qm(D)←Qj(D)；//注释:我们在第三节-第四节中提到了这个案例(2A)。

16: 否则，如果 σm < min(t),则

17: //注释:部分重用旧噪音的情况:

18: 客户计算_平方_成本，使得[高斯(Qm，√_平方_成本，δ 预算) $]^{-2}=\sigma_m{}^{-2}-[\min(\mathbf{\Sigma}_t)]^{-2}$；；

19:客户计算_平方_剩余_预算↓_平方_剩余_预算_平方_成本；20:如果_平方_剩余_预算≥ 0，则

$\sqrt{\sigma_m{}^2-\frac{\sigma_m{}^4}{[\min(\mathbf{\Sigma}_t)]^2}}$ 21: 区块链计算噪声 σm2 和附加噪声← N(0，1) ×$[\min(\mathbf{\Sigma}_t)]^2$

22: 区块链联系服务器进行计算 $(\mathbf{\Sigma}_t)$; // $\widetilde{Q}_{t,\min}(D)[\widetilde{Q}_{t,\min}(D)\,\widetilde{Q}_m(D)$ **2B←Qm(D)+噪声 Qm(D)]+额外噪声，其中表示对应于第三-D 节中最小注释(我们参考此案例)的噪声响应(保持在区块链中)。**

23: 区块链记录 Qm 的查询类型，m，δm，σm， $\widetilde{Q}_m(D)$**;**

24: 其他

25: 返回隐私预算不足的错误；

26: 如果…就会结束

27: 其他

28: //点评:完全复用一个旧噪音的情况:

29:σ 表示 t 中也小于 σm 的最大可能值，区块链重用 Q(D)，表示对应于 σ 的噪声响应(保持在区块链中)；

30: 区块链计算 Qm(D)←Q$(D)+\mathcal{N}(0,1)\times\sqrt{\sigma_m{}^2-\sigma}$ **2；评论:我们参考案例 2C)** ***第三节 d。***

31: 区块链记录 Qm 的查询类型，m，δm，σm，Qm(D)；
32:如果…就会结束
33:结束 if

来自数据集 D 或其相邻数据集 D 不同的输入，达到(ϵ, δ)-DP 如果从 D 的一个记录。
(，δ)-DP 的形式定义在定义 1 中给出，相邻数据集的概念

在注释 2 中讨论。

定义 1((，δ)-DP [8]):一种随机化机制 Y，在给定数据集作为

其中 P[ ]表示概率，概率空间在随机机制 y 的硬币翻转上方。

备注 1:δ= 0 下的(，δ)-DP 的概念变为-DP。-DP 和(，δ)-DP 在许多研究中也被分别称为纯 DP 和近似 DP[9]–[11]。

*备注 2(相邻数据集的概念):两个数据集 D 和 D 如果它们仅在一个元组中不同，则称为相邻的。*关于这一点仍然有不同的说法。在第一种情况下，D 和 D 的大小 相差一，所以 D 是通过向 D 中添加一条记录或从 D 中删除一条记录而获得的。在第二种情况下，D 和 D 具有相同的大小(比如 n)，并且仅在 n 个位置中的一个位置具有不同的记录。最后，相邻数据集的概念也可以定义为包括上述两种情况。我们在本文中的结果适用于上述所有情况。

在实现动态规划的各种机制中，文献[6]中提出的实值查询高斯机制受到了广泛关注。文献[8]给出的改进结果是引理 1.引理 1(德沃克和罗斯[8]的定理 A.1):用 2-灵敏度 Q 回答一个查询 Q，加一个零-√ ×
对于真实查询结果的每个维度，标准差为 2ln(1.25/δ)的平均高斯噪声达到(，δ)-DP。查询 Q 的上述 2-敏感度 Q 被定义为任意两个相邻数据集 D 和 D 的真实查询结果之间的最大 2 距离 不同的记录；即，。*Q =最大相邻 D，D Q(D)-Q(D)2.*

关于查询的 2-敏感性的更多讨论在第三节给出。第七章讨论了隐私参数和 δ 的设置。

B. 区块链、以太网和智能合约

*区块链:区块链技术广泛应用于要求高安全性和高透明度的系统，如比特币和以太币[20]。*通过使用点对点网络，可以有效地利用区块链来解决比特币交易中的双重支出问题。解决方案是在基于哈希的工作证明链(PoW，由比特币使用)中哈希事务信息，这是用于确认事务并向链中产生新块的共识机制算法。一旦记录形成，除了重做 PoW 之外，它不能被更改。

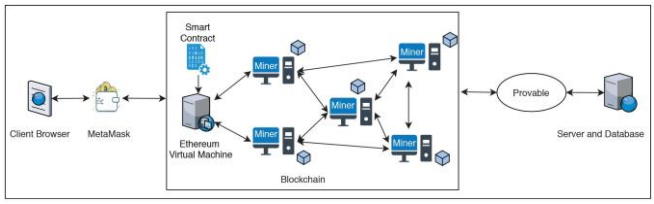
对于 D 和 D 遍历所有相邻数据集对，对于 迭代输出范围的所有子集(1)

Y

此外，区块链还在不断发展，不断增加“完整”的街区。由最近的事务组成的块按时间顺序添加到链中[21]。每个区块链节点都可以有一个区块链的副本。区块链允许参与者在没有集中控制的情况下跟踪他们的交易。

*Ethereum: Ethereum 是一个区块链平台，允许用户创建分散的端到端应用程序[22]。*Ethereum 中的矿工使用 PoW 共识算法来完成事务验证和同步。此外，Ethereum 还可以运行下文详述的智能合约。

*智能合同:智能合同最早是由尼克·萨伯提出的，作为一种可以自动执行合同条款的计算机化交易协议[23]。*它打算



图。1.提出了基于区块链的差异化隐私成本管理系统架构。

以数字方式订立合同，并允许在没有第三方的情况下维持可信的交易。随着区块链的发展，例如以太琴，智能合约作为脚本存储在区块链。一个拥有 Turingcomplete 编程语言的区块链允许每个人为交易定制智能合约脚本[24]。当在区块链上创建或生成完成特定任务或服务的交易时，会触发智能合同。

三. 系统描述

我们基于区块链的系统为查询提供不同的私人响应，同时通过噪声重用最小化隐私成本。我们设计了一个网络应用程序来实现我们的算法 1，该算法通过设置旧的噪声响应的最佳重用比例并添加新的噪声(如果需要)，以最小的隐私成本生成对查询的噪声响应。为了清楚起见，我们将算法 1 及其讨论推迟到第三节。系统的设计如图 1 所示。我们在下面讨论细节。在第五节中，我们将讨论我们基于区块链的系统的实现和实验，并提供更多关于实现的数字。

特别是图。图 3 显示了我们基于区块链的隐私管理系统[25]的截图。4 在使用系统时显示输出。

A. 系统结构

我们的系统包括客户端、区块链、服务器和智能合约，下面是更多细节。

*客户端:客户端的主要功能是将用户的查询转移到区块链智能合约。*客户端使用隐私参数和 δ 计算服务器生成高斯噪声所需的参数标准差，并将查询转发给区块链。此外，在获得对查询的嘈杂响应后，客户端可以向分析师显示查询结果。

*区块链智能合约:区块链充当客户端和服务器之间的中间件。*它决定应该向服务器提交哪个查询。区块链记录剩余的隐私预算、查询类型、回答查询的噪声响应、隐私参数和相应的噪声量。如果剩余的隐私预算足够，智能合约将使用记录的历史执行查询匹配功能。否则，智能合约将拒绝此查询。如果当前查询与历史记录中的任何查询不匹配，智能合约将调用服务器来计算结果。如果之前已经接收到该查询，则如果噪声响应可以完全由旧的噪声回答生成，则区块链智能合约将不调用服务器，并且如果仍然需要访问数据集来生成噪声响应，则将调用服务器。

C. 对手模型

我们系统的对手模型类似于[19]。假设有两种对手。

首先，对手可以获得扰动的查询结果。他们可能会尝试使用受干扰的查询结果来推断用户的真实信息。

其次，对手试图修改隐私预算。例如，他们希望减少使用的隐私预算，这样用户可能会超出隐私预算。这样一来，隐私就会泄露。然而，在我们的案例中，隐私预算记录在区块链上。一旦隐私预算存储在区块链，对手就不能篡改它。

D. 基于重复使用噪声的算法 1

我们在第三章第五节中提出了算法 1 中重复使用噪声的解决方案。我们考虑实值查询，以便可以使用高斯机制。对无值查询的扩展可以看作是未来的工作，我们可以应用[12]的指数机制。

为了澄清符号的用法，我们注意到“奇”指的是“第七次查询”

(按时间顺序排列)并由随机

算法齐。类型 t-query 意味着查询的类型是 t。在不同时间

TABLE II EXAMPLE TO EXPLAIN ALGORITHM 1

Q_m's query type	Q_1=type-1	Q_2=type-2	Q_3=type-3	Q_4=type-1	Q_5=type-2	Q_6=type-1	Q_7=type-3	Q_8=type-2	Q_9=type-2	Q_{10}=type-1	Q_{11}=type-2	Q_{12}=type-1	Q_{13}=type-3
σ_m computed by Line 1 of Alg. 1	$\sigma_1 = 1$	$\sigma_2 = 3$	$\sigma_3 = 2$	$\sigma_4 = 2.5$	$\sigma_5 = 2$	$\sigma_6 = 0.5$	$\sigma_7 = 2$	$\sigma_8 = 2.5$	$\sigma_9 = 1.5$	$\sigma_{10} = 0.25$	$\sigma_{11} = 1$	$\sigma_{12} = 0.75$	$\sigma_{13} = 1.5$
Case involved in Alg. 1	1):$Q_1 \leftarrow Q_1 + \mathcal{N}(0,1)\times\sigma_1$ with accessing D	1):$Q_2 \leftarrow Q_2 + \mathcal{N}(0,1)\times\sigma_2$ with accessing D	1):$Q_3 \leftarrow Q_3 + \mathcal{N}(0,1)\times\sigma_3$ with accessing D	2C): $\widetilde{Q}_4$ reuses $\widetilde{Q}_1$ without accessing D	2B): $\widetilde{Q}_5$ reuses $\widetilde{Q}_2$ with accessing D	2B): $\widetilde{Q}_6$ reuses $\widetilde{Q}_1$ with accessing D	2A): $\widetilde{Q}_7$ reuses $\widetilde{Q}_3$ without accessing D	2C): $\widetilde{Q}_8$ reuses $\widetilde{Q}_5$ without accessing D	2B): $\widetilde{Q}_9$ reuses $\widetilde{Q}_5$ with accessing D	2B): $\widetilde{Q}_{10}$ reuses $\widetilde{Q}_6$ with accessing D	2B): $\widetilde{Q}_{11}$ reuses $\widetilde{Q}_9$ with accessing D	2C): $\widetilde{Q}_{12}$ reuses $\widetilde{Q}_6$ without accessing D	2B): $\widetilde{Q}_{13}$ reuses $\widetilde{Q}_7$ with accessing D

*服务器:数据提供者托管服务器。*服务器提供 API 来回答分析师的查询。当调用应用编程接口时，服务器将查询数据集以计算相应的答案。在计算出真值 Q(D)后，服务器会添加噪音来干扰答案。然后，服务器将嘈杂的答案返回给区块链。

在本文的剩余部分，我们使用区块链、客户端和服务器分别表示区块链、客户端和服务器。

B. 系统功能

*将查询与查询历史记录匹配并生成干扰响应:区块链将当前查询类型与保存的查询类型进行比较，以检索以前的查询结果。*如果是区块链第一次看到查询，区块链会将查询转发给服务器，服务器会将满足 DP 的扰动结果返回给区块链。如果当前查询类型与以前答案的查询类型相匹配，块链会将计算的噪声量与同一查询类型下所有以前保存的噪声量进行比较。根据比较结果，区块链将完全重用旧响应或调用服务器。

管理隐私预算:区块链会在回答查询时更新隐私预算，一旦指定的隐私预算用完，区块链会确保回答查询不会产生新的隐私成本。

询问的查询可以具有相同的查询类型。这就是我们在算法 1 中重用噪声的原因。

, ,Q_{m-1}, where the *i*th query Q_i for i = 1,2,...QM-1，其中 i = 1，2 的第 I 个查询 Qi，...,。**对于 i = 1，2，...,。*m，我们定义 σi :=高斯(Qi，I，δi)，其中 Qi 表示 Qi 的 2-敏感性，这里我们把 Qi 的讨论推迟到第三节-H。***如算法 1 所示，我们有几个案例在下面讨论。为了更好地理解这些情况，我们稍后讨论第二节表二中给出的一个例子。

情况 1):如果第一次看到 Qm，我们通过添加一个具有标准偏差的零均值高斯噪声来获得噪声响应 $\widetilde{Q}_m(D)$

————— 高斯(Qm，m，δm)独立于真实结果 Qm(D)的每个维度(如果隐私预算允许)，如算法 1 第 7 行所示，其中高斯V × (Qm，m，δm) :=

2ln(1.25/δm)(Qm/m)来自引理 1。

t := {σ j: σ j 已记录在 Blockchain 中，Qj 是类型 tquery}。**Blockchain 比较 σm 和 t 中的值，得到以下子案例。**

案例 2A):如果存在 $\sigma_j \in \Sigma_t$ ～，使得 **σm = σj**，那么 **Qm(D)被设置为 Qj(D)。**

案例 2B):该案例认为 σm 小于最小(t)，表示最小

～ (Σ_t) =妈妈在 t。让 Qt，min(D)表示对应于 **min**(Σ_t)的噪声响应(保持在区块链中)；具体来说，如果最小 **σj 为一些 j，那么 Qt，最小(D) = Qj(D)。**下面的

$\sigma_m < \min(\Sigma_t)$，为了降低隐私成本，

我们再用 $[\sigma_m^2/([\min(\Sigma_t)]^2)]$ 产生 **Qm(D)的 Qt、min(D)**中的噪声分数(如果隐私预算允许)。这将通过定理 1 的结果 2)得到，将在 部分三. e。具体来说， 下面的

$\min(\Sigma_t) > \sigma_m$，如第 22 行所示算法 1，**Qm(D)**由$\widetilde{Q}_m(D)$ ←设定

$Q_m(D)$ + [σm2([最小值(t)]2)] × $[\min(\Sigma_t)]^2)\sqrt{\sigma_m^2 - [\sigma}$ **[Qt，最小值(D)-Qm(D)]+N(0，1)×0×**

4/(|. 注意，**m**

如果 **Qm** 是多维的，将根据上述公式在每个维度上添加独立的高斯噪声。这也适用于本文的其他地方。

情况 2C):这种情况认为 **σm** 大于 **min(t)，σm** 与 **t** 中的所有值都不同。设 **σ** 是 **t** 的最大可能值，也就是

小于 **σm**；即。$\sigma_\ell \simeq \max\{\sigma_j : \sigma_j \in$ $(D)+\mathcal{N}(0,1)\times\sqrt{\sigma_m^2}$ σ_ℓ^2 **t 和 σj < σm**}。那么 **Qm(D)被设置为**

Q −. 这将通过定理 1 的结果 2)在第三节变得清楚。

*解释算法 1 的示例:表二提供了一个更好地理解算法 1 的示例。*我们考虑三种类型的查询。特别地，Q1、Q4、Q6、Q10 和 Q12 是类型 1 查询；Q2、Q5、Q8、Q9 和 Q11 是类型 2 查询，Q3、Q7 和 Q13 是类型 3 查询。

E. 解释算法 1 的噪声重用规则

我们的算法 1 的噪声重用规则旨在最小化累积的隐私成本。为了解释这一点，受[13]的启发，我们定义了隐私损失来量化隐私成本。我们分析隐私损失来表征隐私如何以细粒度的方式退化，而不是使用 Kairouz 等人的合成定理。[26]. 虽然[26]给出了差分私有算法组成的最新结果，但这些结果并没有假设实现差分私有算法的基本机制。在我们的分析中，通过具体分析高斯机制的隐私损失，我们可以获得更小的隐私成本。

对于随机化算法 Y，相邻数据集 D 和 D ，并输出 y，即隐私损失 LY(D，D ；Y)表示当随机化算法 Y 应用于 D 和 D 时，观察到相同输出 Y 的概率之间的乘法差 。具体来说，我们定义

$$:= \ln \frac{\mathbb{F}[Y(D) = y]}{\mathbb{F}[\quad' = y]},$$ **LY(D，D ；Y(2)Y(D)**

其中 **F[ ]**表示概率密度函数。

为了简单起见，我们通过假设随机化算法 Y 具有连续输出来使用上面(2)中的概率密度函数 F[ ]。如果 Y 有离散输出，我们用概率质量函数 P[ ]代替 F[ ]。

当 Y 遵循随机变量 Y(D)的概率分布时，LY(D，D ；y)遵循随机变量 LY(D，D)的概率分布 ；Y(D))，我们写为 LY(D，D))为了简单起见。

正整数 m 乘以 Y1Y2 的 Ym. . . Ym。对于合成，相对于相邻数据集 D 和 D 的隐私损失 当随机化机构 Y1、Y2 的输出，...,。*Ym 是 y1，y2，...,。ym 的定义如下*

幺米

$$: = \ln \frac{\overset{i=1}{\quad} \quad i \quad\quad i}{\mathbb{F}\left[\cap_{i=1}^{m}\left[Y_i(D') = y_i\right]\right]}.$$

, ,*m*}, clearly $L_{Y1Y_2\ldots Y_m}(D,D$;$y_1,y_2,\ldots$m}， 显 然 是 LY1Y2...Ym(D，D ；y1，y2，...,。**ym)**遵循随机变量的概率分布

写为 **1 x2...LYYm1(DY2，D...**；YmY(1D(D)，D，y)2 为简单起见。(四)，...,。***Ym(D))，*** ***哪个 我们***

有了上面定义的隐私损失，我们现在分析在 **DP** 下回答一系列查询时如何重用噪声。为此，我们提出了定理 **1**，它给出了重复使用噪声以最小化隐私成本的最佳比率。

定理 1(重用噪声最小化隐私成本的最佳比率):假设在回答查询 Qm 之前

在回答~ ~问题后 ,). 为了 ThD)由 Q1，mQth 查询 2 给出，...,。NQ(m[QA-m(1D，假设，隐私丢失，D)/2]，A(DQ，Dm))是给一些人的

我们在 Qj(D)中重用 r 部分噪声，生成满足 σm2 R2σJ2 > 0 的 0 ≤ r ≤ 1 的 Qm(D),其中 r 是待定常数。

如果 Qj(D)-Qj(D)遵循均值为 0、标准差为 σj 的高斯概率分布，我们生成噪声响应$\widetilde{Q}_m(D)$来回答查询 Qm，如下所示:

$$Qm(D) \leftarrow (D) + r\left[\widetilde{Q}_j(D) - Q_j(D)\right] + \mathcal{N}\left(0, \sigma_m{}^2 - r^2\sigma_j{}^2\right) \quad (3)$$
Qm

因此，Qm(D)-Qm(D)遵循高斯概率分布，均值为 0，标准差为 σm。

注意，Qm 和 Qj 是相同的，因为 Qm 和 Qj 是相同的。然后，我们有以下结果。

1) Qm，

$$D') + [([\|Q_m(D) - Q_m(D')\|_2]^{\sim 2}(1^{\sim} - r)^{\sim 2})/(\sigma_m{}^2 -$$
(Bprivacyr(D，D 失败 LQ1Q))2...forQm(D，BDr()D，德韦尔) :be=))/2]，Br(D，D A(D，r2σj2)]。

2) $\left(\frac{m}{\sigma_j}\right) = \begin{cases} 1, \\ \sigma \quad 2 \end{cases}$ 我们在上述(3)中明确要求 r ≥ 0 且 σm2 R2σJ2≥0[注意 N(0，0) ≡ 0]。 最小化总隐私成本[相当于最小化 Br(D，D)以上]，最优 r 由下式给出

如果 σm ≥ σj，

最佳(4)

，如果 σm < σj，

因此将(4)代入表达式中)给出

Broptimal (D，d)

$$\begin{cases} \text{A(D,D} \quad \text{),} & \text{如果 } \sigma m \geqslant \sigma j \\ = \text{A(D,D} \quad \text{)} + [\|Q_m(D) - Q_m(D')\|_2]^2\left(\frac{1}{\sigma_m{}^2} - \frac{1}{\sigma_j{}^2}\right) & \text{如果 } \sigma m < \sigma j。 \end{cases} \quad (5)$$

m ^ 1 }，我们有 roptimal = 1，只是将 Qm(D)设置为 Qj(D)。

证据:证据在附录 a 中。■

定理 1 的等式(4)清楚地指示了算法 1 中的情况 2B(见算法 1 的第 22 行)的噪声使用比率[σm2/([min(t)])2]，以及算法 1 中的情况 2A)和 C)的噪声使用比率 1(见算法 1 的第 15 行和第 30 行)。

通过考虑定理 1 的结果 1)中的 r = 0，我们得到了推论 1，它给出了高斯机制单次运行的隐私损失的经典结果。

推论 1:通过考虑 m = 结果 1)中的 1

定理 1，对于向查询 Q 添加高斯噪声量 σ 的随机化算法 Q，我们有关于相邻数据集 D 和 D 的隐私损失 由 下 式 给 出)/2]，A(D，D))为 A(D，D):=[(Q(D))

问: $(D')\|_2]^2)/\sigma^2]$.

推论 1 已在许多关于动力定位高斯机制的先前研究[8]–[10]中显示。

通过考虑定理 1 的结果 1)中的 r = 0，我们得到推论 2，推论 2 给出了朴素算法的隐私损失，其中对每个查询的噪声响应是使用新鲜噪声独立生成的。

*DP 下的 Qn。*具体来说，对于 i = 1，2，...,。*n，在(I，δi)-DP 下回答第 I 次查询 Qi，通过将独立的高斯噪声 σi := 高斯(Qi，I，δi)加到真实查询结果 Qi 上，生成噪声响应 Qi，其中 Qi 为 Qi 的 2-灵敏度。*然后，在回答了 Q2 Q1 的 n 个问题后，...,。*如上所述，关于相邻数据集 D 和 D 的隐私损失 是由* $\mathcal{N}([F(D, D')/2], F$ *(D，D))对于 F(D，D)* $:= \sum_{i=1}^{n}([\|Q_i(D) - Q_i(D')\|_2]^2/\sigma_i{}^2)$.

F. 解释算法 1 中的隐私成本更新

在上述情况中，案例 2A)和 C)不会产生额外的隐私成本，因为它们只是使用先前有噪声的结果并产生新的高斯噪声，而不访问数据集 d。相比之下，案例 1)和 2B)会产生额外的隐私成本，因为它们需要访问数据集 D 来计算真实的查询结果 Qm(D)。因此，在算法 1 中，隐私成本在情况 1)和 2B)中被更新，但是在情况 2A)和 C)中不被更新。在本节中，我们将根据案例 1)的第 3 行和第 5 行以及案例 2B)的第 18 行和第 19 行解释算法 1 中更新隐私成本的原因。

$\widetilde{Q}_i$当使用我们的算法 1 时，我们让上面的随机机制 Yi 作为 我们的噪声响应函数。

$\widetilde{Q}_1, \widetilde{Q}_2$

, ,*Qi*–1 on data set*D* are instantiated as ~] 2*y* ,*y* ,...数据集 d 上的 qi 1 被实例化为⎚y，y，...,。 ~ *j* < *yi-1，如果数据*

*集 d 上 Qi 的生成对某些∅. I 使用 Qj，那么 Qi 输入中的辅助信息 auxi 包含 yj (aux1 是。*对于我们算法 1 的连续使用，将变得清楚的是，隐私损失，定义为

ym)

$$\overline{\mathbb{F}\left[\cap_{i=1}^{m}\left[\widetilde{Q}_i(D')=y_i\right]\right]}$$
$$\mathbb{F}\left[\cap_{i=1}^{m}\left[\widetilde{Q}_i(D)=\text{彝族}_i\right.\right.$$

:= ln 最大

相邻数据集

遵循高斯概率分布，对于某些 V，均值为(V/2)，方差为 V，用 N([V/2]，V)表示。由于这种原因，隐私损失的形式，相应的微分隐私级别由以下引理给出。

引理 2:如果随机机制 Y 相对于相邻数据集 D 和 D 的隐私损失 由 N([V(D，D) (ϵ,δ) *2]，V(D，D))对于一些 V(D，D)，那么对于满足最大邻近数据集 D，D 的和δ，Y 达到-DP 五(丁、丁)=[高斯(1，，δ)]2。*

证明:证明详情见附件二。■

基于上面定义的隐私损失，我们有以下定理来解释算法 1 中更新隐私成本的规则。

*定理 2:我们在这里考虑算法 1 的连续使用。*假设回答后 *回答问题前，Qm-1 和 Qm，D 等式由下式给出)/2] ，A(D，D))对于一些 A(D，D)，相应的隐私级别可以由 (o1d，δ budget)DP 给出。*然后，在算法 1 中，在回答了 Q2 Q1 的所有 m 个查询之后，...,。*Qm-1，Qm，我们有如下。*1)相对于相邻数据集 D 和 D 的隐私损失 ：

① 仍将是 N([A(D，D)/2]，A(D，D)在案例 2A)和 C)中)；

② 将是 N([B(D，D)/2]，B(D，D))在情况 1)中为 B(D，D):= A(D，D))+ ([Qm(D))

质量管理)2]2/σm2)；

③ 将是 N([C(D，D)/2]，C(D，D)在 2B 的情况下)

丙(丁，丁):= A(D，D)) + [‖$Q_m(D)$ − Q (D') ‖$_2$]^{2}m ×

[(1 个)(σm2) − $(1/[\min(\Sigma_t)]^2)$.

2)相应的隐私级别可以由下式给出(新的 δ 预算)-具有以下新内容的 DP:

④ 新=旧(在案例 2A)和 C)；

⑤ new2 = old2 + _squared_cost(在案例 1 中)对于_squared_cost 满足

高斯(Δ_{Q_m}:√平方成本，δ 预算)= σm ，

⑥ 新 2 = 旧 2 + _平方_成本 在

情况 2B) 为_平方_成本令人满意的

[高 斯(Qm，√平 方_成 本 ，δ 预 算)]2 =σm $^{-2}$ − [$\min(\Sigma_t)$]$^{-2}$。

定理 2 解释了算法 1 中更新隐私成本的规则。具体而言，结果⑤给出了情况 1)的第 3 行和第 5 行，结果⑥给出了情况 2)的第 18 行和第 19 行
案例 2B)。

证据:证据在附录 c 中。■

G. 总隐私成本分析

基于定理 2，我们现在分析当我们的系统连续调用算法 1 时的总隐私成本。

在没有回答查询的开始，我们有 V = 0(注意 N(0，0) ≡ 0)。然后，通过推论 1 和定理 2 的归纳，对于算法 1 的连续使用，对于某些 V，隐私损失总是以 N([V/2]，V)的形式存在。在我们的算法 1 中，只有当被回答的查询属于情况 1)和 2B)时，隐私损失才改变。更正式地说，我们有以下定理。

, ,Q_n, let N_1, N_{2A}, N_{2B}, and N_{2C} be the set of i ⍰ {1,2,...Qn，设 N1、N2A、N2B 和 N2C 为 I∞{1，2，...,n}这样的齐分别是在例 1))))))中。对于案例 2B 中的查询)，让 T2B 成为查询类型集。 在案例 2B)中，对于查询类型 t ∈ T2B，假设类型 t 查询的数量是 mt，并且让这些类型 t 查询是 Qjt，1，Qjt，2，...,索引 jt，1，jt，2，...,＞σ jt，mt，对于 k∞{ 2，3，...,。*mt},* $Q_{j_{t,k}}$ *～是通过重用* $(\sigma_{j_{t,k}}{}^2/\sigma$ *jt，k 12)来回答的*

旧噪声的分数 in2 乘以 σjt，k jt，k-1 Qqjtj，tk，k-11；更具体地说，]+N(0，σjt，k2 $Q_{j_{t,k}}=Q_{j_{t,k}}+$ [σJT，kωσJT，kω1])

(σjt，k

↓来自案例 2B 第三-E 节算法 1 的第 22 行)。我们还认为，Qjt 是通过在 Qjt，0 中重用(σjt，12/σjt，02)部分旧噪声来解决的.让 type-t 查询的 2-敏感性为(type-t)。

在表二提供的示例中，我们有 N1 = {1，2，3}，N2A = {7}，N2B = {5，6，9，10，11，13}，N2C = {8，12}。

T2B = {type-1，type-2，type-3}。11 ∼2,110 ∼ 1,1在案例 2B)中，类型 1 查询的数量是 m1 = 2，这些类型 1 查询是 Q6 和 Q，因此 j = 6 和 j1，2 = 10(也是 j1，0 = 1，因为 Q6 重用了 Q1)；类型 2 查询的数量是 m2 = 3，这些类型 2 查询是 Q5、Q9 和 Q Q2，因此 j = 5 和 j2，2 = 9，j2，3 = 11(也是 j2，0 = 2，因为 Q5 重用了)；type-3 查询的数量是 m3 = 1，这个 type-3 查询是 Q13，所以 j3，1 = 13(也是 j3，0 = 3，因为 Q13 重用了 Q3)。

然后，在算法 1 被用于回答所有 n 个查询并且查询 Qi 在(I，δi)-DP 下被回答之后，我们有

1)相对于相邻数据集 D 和 D 的总隐私损失 由下式给出)/2]，G(D，D))，在哪里

∑ ‖ − [齐(D) 齐(D) 2]2

G(D，D) := /

σ i2

N1

$$\left\{ \frac{\left[\left\|Q_{j_{t,0}}(D)-Q_{j_{t,0}}(D')\right\|_2\right]^2}{\sigma_{j_{t,0}}{}^2} \right\}$$

σ jt，mt 2

(6)

第一次求和是来自情况 1)中查询的贡献，第二次求和是来自情况 2B)中查询的贡献。当 D 和 D 迭代相邻数据集的空间，Qi(D)的最大值 齐(D)是 齐 的二敏 Qi，以及 Qjt、mt (D)和 Qjt、mt (D) 2 的最大值

Qjt，0(D)-Qjt，0(D) 2 是(type-t)因为 Qjt，mt 和 Qjt，0 都是 type-t 查询，我们得到

最大 总

设计师，总设计师

相邻数据集

$$=\sum_{i\in N_1}\frac{\Delta_{Q_i}{}^2}{\sigma_i{}^2}+\sum_{t\in T_{2B}}\left[\frac{\left[\Delta(\text{type-}t)\right]^2}{\sigma_{j_{t,m_t}}{}^2}-\frac{\left[\Delta(\text{type-}t)\right]^2}{\sigma_{j_{t,0}}{}^2}\right].$$

(7)

在第二节表二提供的示例中，最大相邻数据集 D，D G(D，D)由下式给出

$$\overline{{}^2}+\frac{\Delta_{Q_2}{}^2}{\sigma_2{}^2}+\frac{\Delta_{Q_3}{}^2}{\sigma_3{}^2}+\left[\frac{\overline{\left[\Delta(\text{type-1})\right]^2}}{\sigma_{10}{}^2}-\frac{\overline{\left[\Delta(\text{type-1})\right]^2}}{\sigma_1{}^2}\right]$$

雌三醇环戊醚 ı

σ 12

$$+\left[\frac{\left[\Delta(\text{type-2})\right]^2}{\sigma_{11}{}^2}-\frac{\left[\Delta(\text{type-2})\right]^2}{\sigma_2{}^2}\right]$$

$$+\left[\frac{\overline{\left[\Delta(\text{type-3})\right]^2}}{\sigma_{13}{}^2}-\frac{\overline{\left[\Delta(\text{type-3})\right]^2}}{\sigma_3{}^2}\right]$$

$$=\frac{\left[\Delta(\text{type-1}\quad\text{type-2})+(\text{type-3})\right]^2}{\sigma_{102}\sigma_{112}}+\frac{\left[\Delta\quad\right]^2}{\sigma_{132}}+\frac{\left[\Delta\quad\right]^2}{\sigma_{132}}.$$

2)根据引理 2，我们的算法 1 的总隐私成本可以由(我们的，δ 预算)-DP 给出，以使我们满意

高斯 1，我们的，δ 预算 −2

= 最大 G(D，D) (8)相邻数据集 D，D

或者(ϵ,δ)发展计划 为 任何的 和 δ 令人满意的

[高斯(1，，δ)]2 =最大邻近数据集 D，D G(D，D).

证据:证据在附录 d 中。■

≥ $\overline{=}G(D,D')$ for := $\sum_i^n{}_1([\|Q_i(D)$ −备注 3:定理 3 可以用来理解我们的算法 1 比独立回答 n 个查询的朴素算法产生更少的隐私成本。如推论 2 中所给出的，关于相邻数据集 D 和 D' is given by $\mathcal{N}$ ′ ([F(D，D (D') 2]$^2/\sigma_i{}^2$)2]，F(D，D))为 F(D，D) 齐。

显然，F(D，D)G(D，D)由上面的(6)给出。根据引理 2，朴素算法的隐私成本可以由(朴素，δ 预算)-DP 给出，用于朴素满足[高斯(1，朴素，δ 预算)]-2 =最大邻近数据集 D，D F(D，D)，它与定理 3 中的(8)和中的高斯(，，)表达式一致

引理 1 暗示

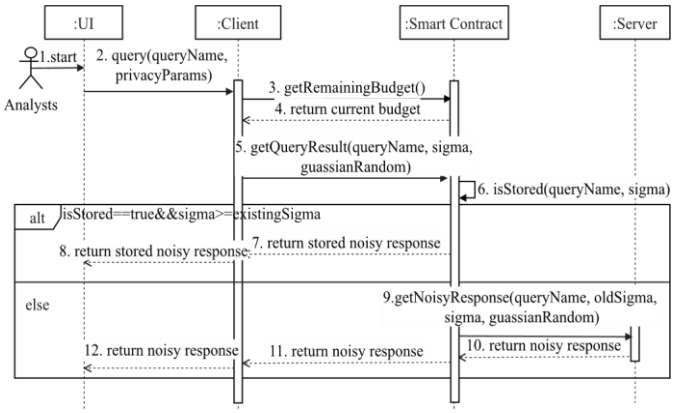
max

我们的= 相邻数据集 G(D，D)≤1,

天真的最大邻近数据集 F(D，D)其中只有当所有 n 个查询都不同时才取等号，因此在我们的算法 1 中不会产生噪声重用。

H. 计算查询的 2-敏感性

查询的 2-敏感性 Q 被定义为任何相邻数据集 D 和 D 的(真)查询结果之间的最大 2 距离，它们在一条记录中不同:Q =最大相邻数据集 D，DQ(D)Q(D)2。对于一维实值查询 Q，Q 只是 Q(D)和 Q(D)之间的最大绝对差)对 于 任何相邻的数据集 D 和 D 。在第五节的性能评估中，我们通过考虑修改条目来定义相邻的数据集。然后，如果数据集有 n 个用户的信息，并且每个用户收入的域在区间[min_income，max_income]内，那么作为所有用户平均收入的查询 Q 的 Q 是[(max _ income-min _ income)/n]，因为这是当用户记录改变时输出的最大变化。类似地，作为女性用户百分比的查询 Q 的 Q 是(1/n)。



图。2.提出了基于区块链的差异化成本管理系统工作流程。

四. 我们基于区块链的系统的实施挑战

我们现在讨论在设计和实施基于区块链的系统过程中的挑战和对策。

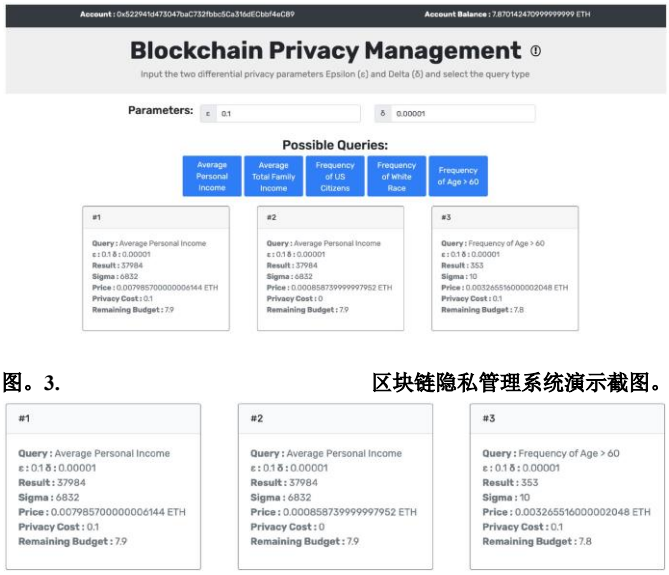
*智能合约获取外部数据:以太网区块链应用程序(如比特币脚本和智能合约)无法访问和直接获取所需的外部数据。*然而，在我们的应用程序中，区块链需要从服务器获取数据，然后将它们返回给客户端。这需要智能合约来发送 HTTP POST 请求。因此，我们使用了 Provable，这是一种集成了许多区块链协议的服务，非锁链应用程序也可以访问它。它保证从原始数据源提取的数据是真实的，没有篡改。

通过使用可证明，智能合约可以直接从网站或应用编程接口访问数据。在我们的例子中，区块链可以用参数向服务器发送 HTTP 请求，然后在服务器成功响应后处理和存储数据。

*具有实体性的数学运算:区块链是使用实体性语言编写的，它是针对以太网虚拟机而设计的。*然而，目前的实度语言对于复杂的数学运算没有固有的功能，例如求平方根或对数。我们编写一个函数来实现平方根运算。为了避免在区块链中使用引理 1 计算对数，我们在客户端生成高斯噪声，并将该值作为函数 QueryMatch 中的参数之一传递给区块链。此外，当前的坚实版本不能操作浮点或双类型数据。为了保持精度，我们在计算过程中放大噪声量，然后在将噪声数据返回给分析师之前缩小该值。

动词 （verb 的缩写）。实现和实验

在这一部分，我们通过实验验证了所提出的系统和算法能够有效地节省隐私成本。2.更具体地说，用户通过用户界面发送查询，然后客户端接收该查询并将其转发给区块链智能合约。在智能合约检查完存储的



图。3. 区块链隐私管理系统演示截图。

图。4. 显示带有隐私成本的输出。

数据，它将决定是直接将有噪声的响应返回给客户机，还是将请求转发给服务器。如果服务器收到请求，它将生成并返回一个对智能合约的嘈杂响应。

A. 实验设置

我们基于第三节中的系统描述，构建了一个网络应用程序的原型。我们使用 Javascript 语言编写 Client，而 Solidity 语言用于区块链智能合约。此外，使用 Web3 作为 Javascript API，在客户端和区块链智能合约之间交换信息，然后利用 Node.js 和 express Web 框架设置 Server。此外，MongoDB 被用作数据库来托管真实世界的数据集。

我们设计的智能合同部署在 Ropsten [27] testnet 上，带有 Chrome 浏览器的元掩码扩展。Ropsten testnet 是一个由 Ethereum 维护的测试区块链环境，它实现了与主 Ethereum 网络相同的 PoW 协议。图。图 3 显示了我们位于区块链的隐私管理系统的截图。图。4 在使用系统发送查询时显示输出。

我们基于包含美国社区调查样本的真实数据集来评估提议的数据保护机制的性能，这些样本是从 https://www.ipums.org 的综合公共使用微观数据系列中提取的。数据集中有 5000 条记录。每个记录包括以下数字属性:“个人总收入”、“家庭总收入”、“年龄”和分类属性:“种族”、“公民身份”。我们将隐私预算设置为预算= 8 和 δ 预算= 10-4，它们通常用于保护数据集的隐私[28]，[29]。我们考虑五种类型的查询:1)“平均个人收入”；2)“平均家庭总收入；”3)“美国公民的频率；”4)“白种人的频率；”和 5)“年龄超过 60 岁的频率.”对于每个查询 Qi 的隐私参数，我们从[0.1，1.1]中均匀采样 I，从[105，104]中均匀采样 δi。这些查询的敏感度分别为 202、404、0.0002、0.0002 和 0.0002。我们根据第三节计算查询的敏感度。对于查询“平均个人总收入”，由于在上述数据集中，用户的个人总收入在 5000 到 700000 之间，我们假设所有可能的数据集中，个人总收入的范围在[-10000，1000000]之间。灵敏度为(1000000(10000))/5000 = 202，该机制保护[-10000，1000000]内所有数据的隐私。因此，它可以保护我们实验中数据集的隐私。假设收到的查询是“平均家庭总收入”。在这种情况下，我们假设所有可能的数据集的最大变化为[-20000，2000000]，因为在我们使用的数据集中，家庭总收入的范围为[-5000，1379500]。灵敏度为(2000000(20000))/5000 = 404。

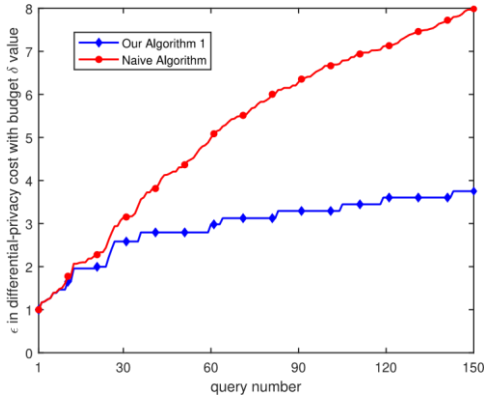
因此，我们产生的灵敏度为 404 的噪声可以保护[20000，2000000]范围内所有数据的隐私。因此，它也可以保护我们使用的数据集的隐私。对查询“美国公民的频率”、“白人的频率”和“年龄超过 60 岁的频率”的敏感度为 1/5000 = 0.0002。

B. 实验结果

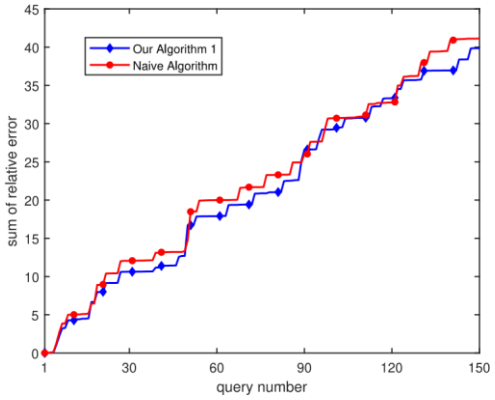
我们实验的基准是一个简单的方案，在智能合约中不包含算法 1。也就是说，每个查询都将由智能合约转发给服务器，以获得嘈杂的响应。因此，在原始方案中，没有 DP 成本可以重复使用。

首先，我们使用一个实验来验证我们提出的算法 1 在节省隐私成本方面是有效的。因此，我们设计了一个性能比较实验，分别使用我们的算法 1 和朴素方案跟踪隐私成本。具体来说，我们在 Ropsten testnet 上部署了两个智能合约，分别实现了我们的算法 1 和朴素方案。然后，我们从 Web 应用程序的 Client 中随机发送五种查询类型中选择的 150 个请求，并记录每个查询的隐私成本。如图 2 所示。5、与朴素方案相比，该算法节省了显著的隐私成本。当查询数为 150 时，算法 1 的差分隐私代价比朴素算法低 52%左右。我们还观察到，当查询数量增加时，所提出的方案中的隐私成本缓慢增加，甚至趋向于收敛到特定值。原因是，在算法 1 中，对于每个查询类型，当查询类型被第二次或更多次询问时，我们总是可以部分或完全重用先前的有噪声的答案。因此，在我们的方案中，如果有噪声的响应完全重用先前有噪声的答案，则许多查询都会得到回答，而不会导致额外的隐私成本。

其次，为了证明所提出的算法 1 保留了数据集的准确性，我们设计了另一个实验来比较相对误差的总和。我们使用与上次实验中相同的智能合约。我们在每个查询中积累相对误差。图。6 表明算法 1 的相对误差之和与朴素算法相当



图。5. 隐私成本总和的性能比较。



图。6. 相对误差和的性能比较。

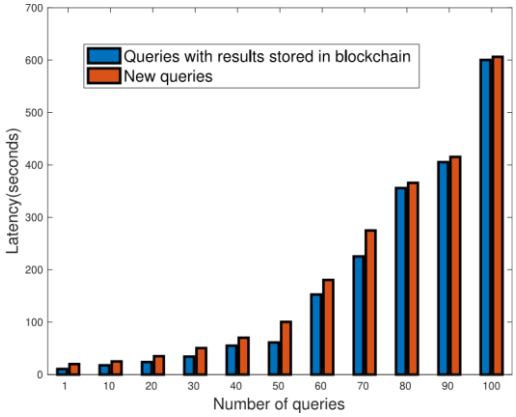
方案。由于两种方案之间的相对误差相似，我们的结果表明所提出的算法 1 保持了精度。

总的来说，图。图 5 和图 6 共同证明了我们的算法 1 可以显著节省隐私成本，而不会牺牲数据集的准确性。

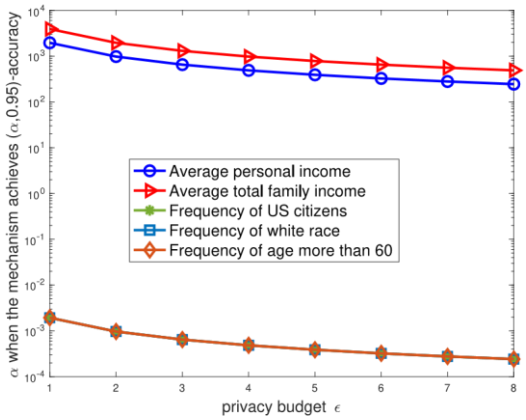
第三，我们评估系统的延迟。我们系统的延迟受区块链、元掩码、网络条件和服务器的影响。为了缩短网络传输速度，我们使用 Ganache-cli 客户端在 http://localhost:3000/ 设置了一个本地 testnet，blockTime 设置为 15 s [30]。图。图 7 显示延迟随着查询数量的增加而增加。以太网的吞吐

量为每秒 20∞60 次传输(TPS)[31]–[36]。当查询数量达到 60 时，延迟会显著增加。除了以太网的吞吐量，元掩码和部署设备的容量都会影响延迟。当查询结果已经保存到系统中时，我们测试这个案例。因此，智能合约不需要向服务器发送请求。我们可以通过使用以前的查询结果来获得查询结果。最糟糕的情况是，智能合约必须为每个查询发送请求，这需要更长的时间才能获得结果，因为第三方服务是可证明的。

第四，我们评估了查询效用和隐私预算之间的关系。如[37]中所定义的，当$|\widetilde{Q}_m(D)$ Qm(D)|≤α 且概率至少为 1β 时，机制的隐私效用满足(α，β)-有用。因此，较小的 α 意味着



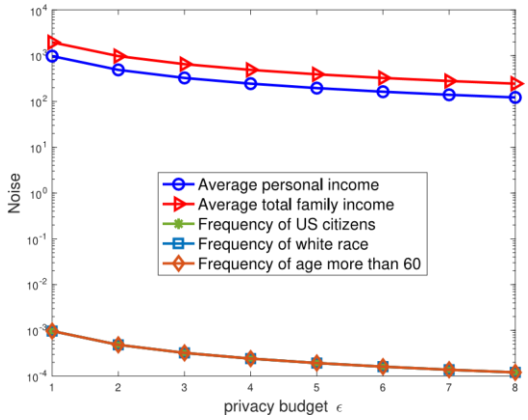
图。7. 延迟与查询数量的关系。



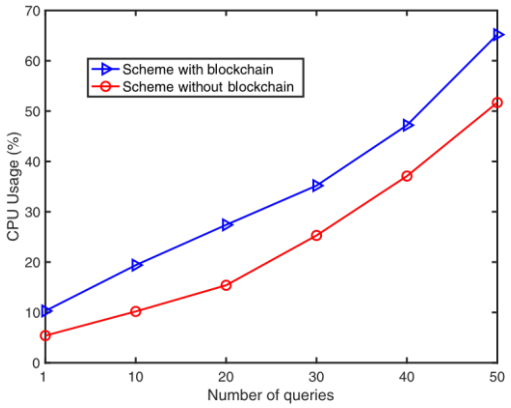
图。8. 效用与隐私预算。

扰动结果与实际结果之间的差异很小，这也反映了该机制具有很高的实用性。添加到查询中的噪声可以计算为√ σ =高斯(Q，，δ)，其中高斯(Q，，δ) = 2ln(1.25/δ) × (Q/)。我们设置 δ= 10 ^ 5 和∈ [1，8]。附录 E 证明了当我们设置 β = 0.05 时，α = 2σ。无花果。图 8 和图 9 说明了效用和噪声如何随着隐私预算的增加而变化。图。图 8 显示了当隐私预算增加时，α 的值降低，这意味着效用增加。此外，添加的噪声量也反映了查询工具。当向查询响应添加的噪声越少，响应获得的效用就越大。图。图 9 显示了噪声如何随着隐私预算而变化。随着隐私预算的增加，噪声会降低，这意味着查询实用程序会增加。噪声的大小取决于隐私预算和敏感度值。诸如“美国公民的频率”、“白人种族的频率”和“年龄超过 60 岁的频率”等查询具有相同的敏感度值 0.0002，因此当它们使用的隐私预算相等时，添加到它们的响应中的噪声是相同的。

第五，我们评估了使用或不使用区块链的 DP 方案所使用的计算资源的效率。没有区块链参与的方案是，数据分析师直接给服务器打电话，获取嘈杂的答案。我们使用“Apache JMeter”[38]应用编程接口测试软件来模拟这种情况，以向



图。9. 噪音与隐私预算。



图。10.比较有或没有智能合约的方案的 CPU 使用情况。

直接服务器。我们将 CPU 使用率视为评估计算资源效率的指标[39]，[40]。

*实验设置:我们使用搭载 2.3 GHz 四核英特尔酷睿 i5 和 8gb 2133 MHz lpddr 3 的 MacBook Pro 来运行应用。*在实验中，我们使用带有三个节点挖掘的“Go-Ethereum”平台部署了一个本地私有的区块链测试网。然后，我们在本地保留一个单独的服务器，用于测试处理从“Apache JMeter”发送的应用编程接口请求的 CPU 使用情况。我们随机发送 50 个查询进行测试。同时，我们使用“活动监视器”软件[41]来跟踪和获取他们的 CPU 使用情况[42]。

*实验结果:图。*10 比较有和没有区块链的方案所花费的 CPU 使用。我们可以观察到计算效率(即，。在我们基于区块链的方案中，CPU 使用率)高于使用服务器直接处理。

不过，还是可以接受的。我们用来构建 Web 服务器应用程序的 Node.js 和 Express.js Web 应用程序框架是 CPU 密集型的。使用区块链时节省计算资源的一种方法是在没有任何传入任务时停止挖掘。只有在必要的时候才开始挖掘。因此，我们提出的方案是有效和实用的，计算成本可以接受。

不及物动词。相关著作

在本节中，我们首先比较我们的论文和一项密切相关的研究[19]，然后讨论其他相关工作。

A. 与杨等比较。[19]

杨等。[19]利用区块链和 DP 技术实现数据共享过程中的安全和隐私保护。与[19]相比，我们将我们的工作与[19]之间的差异总结如下。

1) 虽然[19，算法 1]声称满足-DP，但它并不满足-DP，因为噪声输出的域(即。所有可能值的集合)取决于输入。解释如下。在[19]中，对于两个相邻的数据集

D 和 D ，存在输出的子集 Y，使得

/0。这意味着 e，也就是

$[(\mathbb{P}[Q(D) \in \mathcal{Y}])/(\mathbb{P}[Q(D') \in \mathcal{Y}])] = \infty >$违反-任何 $<\infty$的差异隐私。

2) 杨等。[19]没有讨论如何在其算法 1 中选择小的附加隐私参数。
3) 在[19]中，当第一次询问查询时，使用[7]的拉普拉斯机制 for -differential privacy 将拉普拉斯噪声添加到真实的查询结果中。事后，杨等人。[19]在先前的噪声输出上增加了新的拉普拉斯噪声，这使得新的噪声响应不再遵循拉普拉斯分布，因为独立拉普拉斯随机变量的和不遵循拉普拉斯分布。因此[19]中的分析是无效的。

我们使用高斯噪声来考虑(，δ)-DP。高斯噪声相对于拉普拉斯噪声的优势在于，由于独立高斯随机变量的和仍然遵循高斯分布，而独立拉普拉斯随机变量的和不服从拉普拉斯分布，因此不同隐私保护算法的组成更容易进行隐私分析。

B. 其他相关工作

DP 作为一种强有力的保证数据库隐私的数学模型，近年来备受关注。区块链是一种快速发展的技术，以分散的方式提供安全性和隐私[18]，[43]–[47]。冯等。[48]总结了以前关于区块链系统隐私保护的研究，包括身份和交易隐私保护的方法。在下文中，我们将介绍更多利用区块链或隐私技术在身份、数据和交易中提供隐私或安全保护的最新研究。

*利用区块链进行身份隐私/安全保护:一些研究侧重于利用区块链来保证访问控制管理或身份保护中的隐私/安全。*例如，Zyskind 等人。[49]和夏等。[50]两者都在访问控制管理中使用了区块链。Zyskind 等人。[49]创建了一个分散的个人数据管理系统，以解决用户在使用第三方移动平台时对隐私的担忧。夏等。[50]提出了一个允许的基于区块链的数据共享框架，只允许经过验证的用户访问云数据。陆等。[51]开发了一个私有且匿名的分散式众包系统 ZebraLancer，该系统克服了传统分散式众包中的数据泄露和身份泄露问题。上述研究集中于身份隐私，因为区块链是匿名的，而没有考虑对数据库的隐私保护。

利用进行数据隐私/安全保护:除了身份隐私保护，胡等。[52]用智能合约取代了中央服务器，并构建了一个分散的隐私保护搜索方案，用于计算加密数据，同时确保数据的隐私，以防止恶意中央服务器的不当行为。Luongo 和 Pon [53]使用安全多方计算设计了一个名为 Keep 的隐私原语，该原语允许合同管理和使用私有数据，而不会将数据暴露给公众区块链用于保护公共区块链上的智能合同。或者，我们使用 DP 标准来保证隐私。此外，区块链流行用于物联网场景中数据共享的安全保护[44]，[45]，[50]。

*利用区块链进行交易隐私/安全保护:此外，以前的一些研究使用区块链来保证交易中的安全和隐私。*比如亨利等人。[17]提议区块链应使用搭载在覆盖网络上的机制，覆盖网络可随时宣布交易，以解除用户网络级信息的链接，而不是使用 Tor 等外部服务来保护用户隐私。Gervais [31]提出了一个定量框架来分析区块链工作证明的安全性，该框架的输入包括安全性、共识和网络参数。Herrera-Joancomartí 和 Pérez-Solà [54]专注于比特币交易中的隐私问题。Sani 等人。[55]提出了一种具有高性能和可扩展性的新区块链 Xyreum，以确保工业物联网中的交易安全。

*利用差异隐私保护物联网数据隐私:物联网设备定期收集用户的使用状态，其中可能包含敏感信息，如能耗或位置信息。*为了避免隐私泄露，一些研究使用数据保护机制来保护数据隐私[1]–[5]。例如，都铎等人。[1]提出了一个基于流的框架 Bes 来公开物联网数据。哈桑等人。[2]将每个物联网节点视为区块链的一个节点，以保证物联网节点的安全性，并利用动态规划机制保护每个节点的数据隐私。为了防止对手拦截来自/去往智能家庭网关的互联网流量，或者通过数字踪迹来描述居民的行为，刘等人。[4]利用 DP 开发一个框架来防止攻击。但是，它们都没有讨论如何重用 DP 预算。

*差分隐私算法:一些差分隐私算法被提出来提供隐私保护。*肖等。[56]提出了一种算法来关联添加到不同查询的 Lapalce 噪声，以提高整体准确性。在给定一系列计数查询的情况下，李和 Miklau [57]提出的机制选择了一个查询子集来私下回答，并使用它们的有噪声的答案来导出其余查询的答案。对于一组不重叠的计数查询，Kellaris 和 Papadopoulos [58]通过精心分组对计数进行预处理，并通过平均对它们进行平滑，以降低灵敏度，从而降低注入的噪声量。考虑到查询的工作量，Yaroslavtsev 等人。[59]

引入了一种解决方案，通过比其他人更准确地回答一些查询来平衡准确性和效率。

七. 讨论和未来的工作

在本节中，我们将讨论 DP 参数的含义、智能合约的隐私以及查询。

A. 差分隐私参数

DP 参数的值表示由 DP 机制提供的保护级别。McSherry 和 Mahajan [60]和 Aaby 等人。[61]DP 的强度量化如下。当= 0 时，DP 机制提供了完美的隐私。那么，当≤ 0.1 时，保护被认为是强的，而≥ 10 时，隐私保护是弱的。隐私参数 δ 代表数据库中记录被更改的小概率，因此它应该非常小。对于每个查询，我们从[105，104]中统一采样 δ。

B. 智能合同的隐私

当智能合约部署到公共区块链时，它是公开可用的。在我们的实验中，攻击者可以获得在智能合约中实现的算法。然而，即使他们获得了算法，他们仍然不能从有噪声的结果中获得准确的响应。有一些方法可以用来保护智能合约的隐私，如下所示:

首先，科斯巴等人。[62]提出了 Hawk，一个构建隐私保护智能合约的框架。Hawk 使程序员能够在不考虑密码学的情况下编写私有智能合约。该框架将在编译过程中自动生成一个加密协议。

其次，我们可以通过将 Ethereum 部署到私有的区块链来部分解决这个问题。我们可以结合私人区块链和权威证明[63]共识机制。当以太网被部署到私有的区块链时，私有的区块链可以设置访问控制。因此，攻击者需要在访问智能合约之前破坏访问控制。因此，当使用私人区块链时，我们考虑访问控制来保护智能合约的隐私。

由于保护智能合同的隐私很复杂，我们希望将智能合同的隐私视为我们未来的工作。

C. 查询隐私

DP 机制认为数据分析师是不可信的，持有数据库的馆长是可信的。受信任的策展人存储数据库，并响应不受信任的分析师的统计查询，这样 DP 就不会保护数据分析师查询的隐私。此外，DP 支持可能不包含太多敏感信息的统计查询。当我们使用智能合约时，一些数据分析师可能会担心他们查询的隐私性。有两种方法可以保护查询的隐私。

首先，我们可以使用私人区块链，并使用私人区块链和权威证明共识机制进行实验。Ethereum 还支持将智能合约部署到私有的区块链。在这种情况下，可以认为智能合约是可信的，这样查询的敏感信息就不会泄露。

其次，我们可以将其他密码技术与 DP 结合起来。比如 Agarwal 等人。[64]设计了支持不同私人统计查询的加密数据库。在他们的论文中，策展人和分析师都是不可信的。管理员可以通过利用 EDb 加密操作，将数据库安全地外包给不受信任的服务器。由于对查询的隐私保护是复杂的，并且可能涉及更多的隐私和安全技术，我们希望将查询的隐私视为我们未来的工作。

第三，通过固定查询类型可以避免查询隐私。由于隐私预算相当有限，不可能让数据分析师问太多问题。因此，一种解决方案是控制查询类型。系统构建者可能需要一些时间来选择数据分析师常用的问题，然后他们设置一个下拉列表供数据分析师选择问题。在这种情况下，我们的系统不会泄露查询的隐私，因为查询是标准的。

D. 我们提出的方案与常规动力定位方案的区别

我们预定义了查询，以有效地计算灵敏度值并节省用户的时间。即使我们不使用区块链，也应该预先定义不同查询的敏感度值。当一个查询以不同的 DP 参数多次出现时，我们的方案将对节省 DP 预算起到至关重要的作用。例如，由于相似的数据分析任务，许多公司试图向数据集发送相同的查询。在这种情况下，可以节省一些隐私预算。由于隐私预算是关于数据集的稀缺资源，因此有必要使用我们的方案。

然而，当第一次看到查询时，我们的方案只能将其视为新查询。由于开发计划署的预算十分有限，敏感性必须事先计算，因此一个数据集将不支持太多不同的统计查询。如果我们的系统在现实世界中实现，类似于图。3、可以提供一个支持的查询列表来控制查询的变化，而不是让数据分析师随意输入不同的查询。

八. 结论

在本文中，我们使用基于区块链的方法来跟踪和节省差异隐私成本。在我们的设计中，我们提出了一种算法，该算法对相同查询类型的不同实例完全或部分地重用噪声，以最小化累积的隐私成本。通过严格的数学证明，证明了算法的有效性。此外，我们设计了一个基于区块链的系统进行现实世界的实验，以证实所提出的方法的有效性。

附录一定理 1 的证明

$\mathcal{N}([A(D,D')/2],A$ QM-1，隐私损失 LQ(D，D)是 由 ~ ~给出的(D，D))对于一些 A(D，D). 稍后，我们将展示这样一个 A(D，D)的存在). 然后，当 yi 遵循随机变量 Qi(D)对于每个 I∞{ 1，2，...,。*m1 }，我们有以下隐私损失* $L\tilde{Q}\,\|\tilde{Q}\,\|\ldots\|\tilde{Q}_{m-1}$ *1 2 (D，D ；y1，y2，...,。ym 1):*

$ym^{-1}\big)$

$$= \ln \frac{\mathbb{F}\left[\cap_{i=1}^{m-1}\left[\widetilde{Q}_i(D) = y_i\right]\right]}{\mathbb{F}\left[\cap_{i=1}^{m-1}\left[\widetilde{Q}_i(D') = y_i\right]\right]} \sim \mathcal{N}\left(\begin{matrix} A(D,D') \\ 2 \end{matrix}, A(')\right)$$

: ，

(9)

我们用“∨”来表示“服从分布”。

Qm。我们看隐私损失 $L\widetilde{Q}_1\ \widetilde{Q}_2\ \cdots\ \widetilde{Q}_m$ (D，D；y1，y2，...,。*ym)定义如下:*

幺米

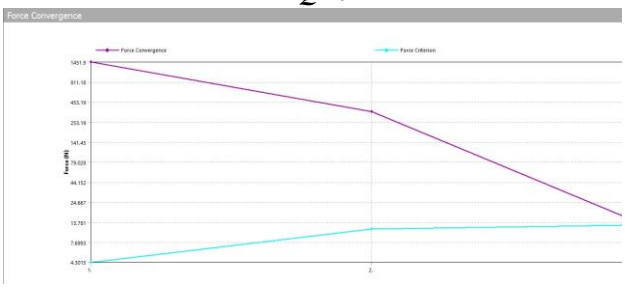
$$: \overline{\mathbb{F}\left[\cap_{i=1}^{m}\left[Q_i(D') = y_i\right]\right]}\mathbb{F}\left[\cap_{i=1}^{m}\left[\underset{\sim}{Q_i}(D) = y_i\right]\right]$$

$$= \ln \qquad \text{。(10)}$$

因此，我们用(9)来分析(10)。从(3)开始，由于$\widetilde{Q}_m(D)$是通过重复使用$\widetilde{Q}_j(D)$和产生附加噪声(如果需要)而产生的，其中 j 是{1，2，...,。*m1}如定理 1 所述，我们有*

$$_1\|\widetilde{Q}_2\|\ldots\|\widetilde{Q}_m\left(D, D'; y_1, y_2, \ldots, y\right.$$

$$\sim \qquad \sim \boldsymbol{LQ\ m}$$



$$\text{ym} \mid_{\boldsymbol{m}} \widetilde{Q}_j(D) = y_j\Big]$$

$$\boldsymbol{ym} \mid \widetilde{Q}_j(D') = y_j\Big]$$

D 品牌

_____ /. **(11)**米 品牌

我们现在讨论$(\mathbb{F}[\cap_{i=1}^{m-}[Q_i(D)$ =的第一个学期 **yi]]= yi]])**和 第 二 项 $\ln[\sim$ **(F[Qm(D)= ym | Qj(D)= yj** $)/(\mathbb{F}[\widetilde{Q}(D')]$**m = y**$|\ Q_j(D') =$ **myj]))]**在**(11)**的最后一行。首先，从**(9)**开始，**(11)**最后一行的第一项遵循高斯分布**)/2]，A(D，D))**. 接下来，我们分析**(11)**最后一行的第二项。

当$\widetilde{Q}_j(D)$ and $\widetilde{Q}_m(D)$分别取 yj 和 ym 时，Qj(D)-Qj(D)和

Qm(D)-Qm(D)r[Qj(D)-Qj(D)]分别取以下定义的 gj 和 gm:

gj:= yj-Qj(D)， (12)

通用汽车:$= y_m - Q_m(D) - r\left[y_j - Q_j(D)\right]$。**(13)**

对于 D 作为 D 的相邻数据集，我们进一步定义

HJ:= Qj(D)Qj(D'), (14)

$= Q_m(D) - Q_m(D')$, (15)

以便

gj+HJ = yj Qj(D'), (16)

通用汽车

$+ h_m - rh_j = y_m - Q_m(D') - r\left[y_j - Q_j(D')\right]$. (17)公

司

注意 **hj** 和 **hm** 是相同的，因为 **Qj** 和 **Qm** 是相同的。通过以上分析，我们得出

$$\mathbb{F}\left[\widetilde{Q}_m(\ ) = y_m \mid \widetilde{Q}_j(D) = y_j\right]$$

$$\left[\widetilde{Q}\ (D) - Q\ (D) - r\left[\widetilde{Q}\ (D) - Q\ (D)\right] = g\right]$$

丁

$$\overset{(b)}{=} \frac{1}{\sqrt{2\pi\left(\sigma_m{}^2 - r^2\sigma_j{}^2\right)}} \cdot e^{-\frac{gm^2}{2\left(\sigma_m{}^2 - r^2\sigma_j{}^2\right)}}, \quad (18)$$

因为$\widetilde{Q}_j(D)$ − **Qj(D)**是一个

方差 **σj2** 和的零均值高斯随机变量

Qm/ Qj(D)]是一个零均值高斯随机变量，方差 σm2-R2σJ2。

同样，对于数据集 **D** ，我们有

$$\mathbb{F}\left[\widetilde{Q}_m(D') = y_m \mid \widetilde{Q}_j(D') = \right.$$

yj*j*

$$\overset{(b)}{=} \frac{1}{\sqrt{2\pi\left(\sigma_m{}^2 - r^2\sigma_j{}^2\right)}} \cdot e^{-\frac{\left(g_m + h_m - rh_j\right)^2}{2\left(\sigma_m{}^2 - r^2\sigma_j{}^2\right)}}, \quad (19)$$

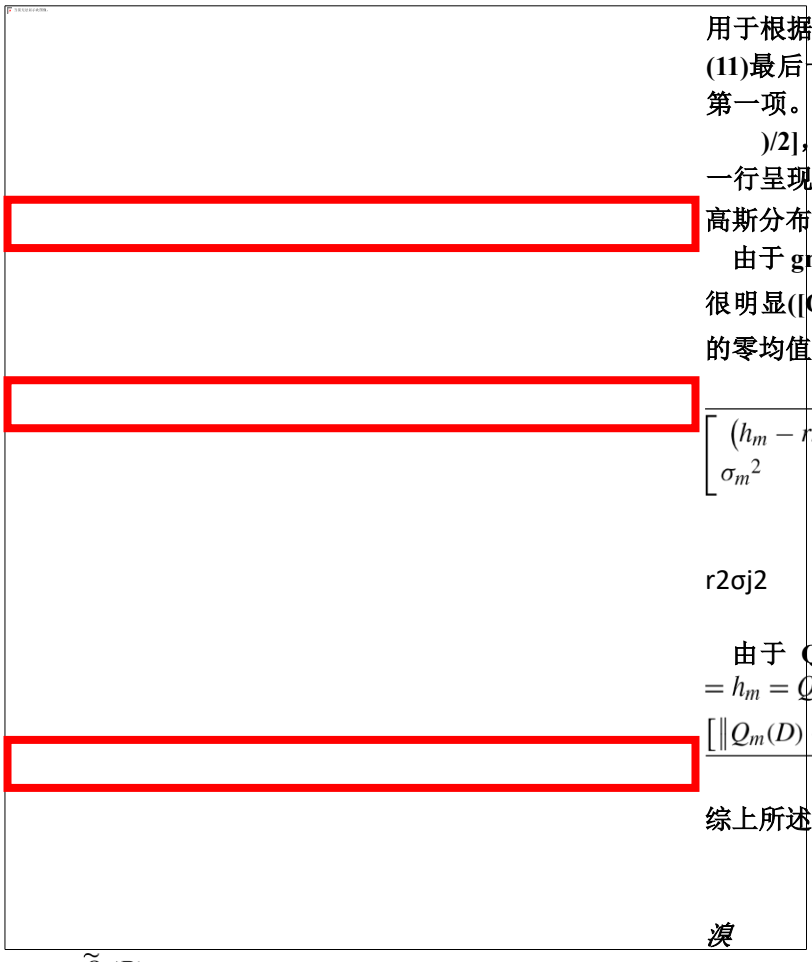
其中步骤**(b)**从 **Qj(D)Qj(D)**是具有方差 **σj2** 和的高斯随机变量

质量管理 $)-Q_m(D')-r[\widetilde{Q}_j(D')-$ ***Qj(D)***

存在 a 零均值

方差为 **σm2 R2σJ2** 的高斯随机变量。

然后



用于根据(3)和(13)生成$\widetilde{Q}_m(D)$的额外高斯噪声的量；即，。(11)最后一行中的第二项实际上独立于(11)最后一行中的第一项。从(9)开始，(11)最后一行的第一项遵循高斯分布)/2]， A(D，D)). 接下来，我们表明(20)在(11)的最后一行呈现第二项也遵循
高斯分布。

由于 gm 遵循方差 σm2-R2σJ2 的零均值高斯分布，因此很明显([GM(hm-rhj)]/[σm2-R2σJ2])遵循方差由下式给出的零均值高斯分布:

$$\left[\frac{(h_m - rh_j)}{\sigma_m{}^2}\right] \times \left(\sigma_m{}^2 - r^2\sigma_j{}^2\right) = \frac{(h_m - rh_j)}{r^2\sigma_j{}^2}. \quad (21)$$

2 2 2

r2σj2 σm2

由于 Qm 和 Qj 是相同的，我们从(14)和(15)得到 hj $= h_m = Q_m(D) - Q_m(D')$，我们用它写(21)为

$$\frac{\left[\left\|Q_m(D) - Q_m(D')\right\|_2\right]^2 (1-r)^2}{\sigma_m{}^2 - r^2\sigma_j{}^2}. \quad (22)$$

综上所述，隐私损失是

[‖ Qr)2²

2

溴 *m* *j*

$\mathcal{N}([A(D,D')/2], A$ QM-1，隐私损失 LQ(D，D) 是 由 ~ ~给出的(D，D))对于一些 A(D，D). 有了上面的结果(23)，我们实际上可以证明确实存在这样的 A(D，D). 这源于数学归纳法。对于基本情况；即，。当只回答一个查询时，结果来自[65，Lemma 3]。诱导步骤由上述结果(23)给出。因此，我们证明了 A(D，D)的存在). 利用这个结果和(23)，我们完成了定理 1 的结果(1)的证明。

2)通过最小化 Br(D，D)得到最优 r)并因此最小化[(1 r)2/(σm2 R2σJ2)]。通过分析这个表达式的单调性，我们得到了(4)中的最优 r。Br(D，D)的一阶导数)到 r 是

$$\mathrm{Br}(D, D') = \frac{-2\left(r\sigma_j{}^2 - \sigma_m{}^2\right)(r-1)}{\left(r^2\sigma_j{}^2 - \sigma_m{}^2\right)^2}. \quad (24)$$

1) *情况 1: if σm* $\geq \sigma j$, $Br(D, D')' \geq 0$ when $r \in$ [1，(σm/σj)]和 Br(D，
D) 当 r∞(∞，1)时< 0 ([σm/σj]，+∞)。因此，使 Br(D，D)最小的最优 r)处于 r = 1。

ym |$\widetilde{Q}_j(D) =$ yj

$$\mathbb{F}\left[\widetilde{Q}_m(D') = \mathbf{ym} \mid \mathbf{Q}(D') = y_j\right]\mathbf{j}$$

$$= \ln \frac{\frac{1}{\sqrt{2\pi\left(\sigma_m{}^2 - r^2\sigma_j{}^2\right)}} \cdot e^{-\frac{g_m{}^2}{2\left(\sigma_m{}^2 - r^2\sigma_j{}^2\right)}}}{\frac{1}{\sqrt{2\pi\left(\sigma_m{}^2 - r^2\sigma_j{}^2\right)}} e^{-\frac{\left(g_m + h_m - rh_j\right)^2}{2\left(\sigma_m{}^2 - r^2\sigma_j{}^2\right)}}}$$

$$= \frac{g_m\left(h_m - rh_j\right)}{\sigma_m{}^2 - r^2\sigma_j{}^2} + \frac{\left(h_m - rh_j\right)^2 - g_m{}^2}{2\left(\sigma_m{}^2 - r^2\sigma_j{}^2\right)}$$

克

= /

2

/. (20)

2 j

上面的(20)给出了(11)最后一行的第二项。乍一看，似乎第一个学期

$\ln\left[\left(\mathbb{F}\left[\cap_i^{m-1}\left[\widetilde{Q}(D) = y\right]\right]\right)/\left(\mathbb{F}\left[\cap^{m-1}\left[\widetilde{Q}(D') = y\right]\right]\right)\right]$和第二项 ln[= | = m(D)= ym | Qj(D)= yj])]在(11)的 最后一行是从属的，因为它们都涉及 yj。然而，我们已经从上面的(20)中表明，(11)的最后一行中的第二项仅依赖于随机变量 gm[注意(20)中除 GM 之外的项都给出了]，这是

2) ***情况 2:如果 σm < σj，Br(D，D)当 r∞[(σm，σj)，1] 和 Br(D，D)时，′ ≥为 0) 当 r∞时< 0(∞，[σm manyσj])∞(1，+∞)。*** 因此，使 **Br(D，D)最小的最优 r)处于 r = ([σm/σj])2.因此，我们得到 r 的最佳值为(4)。**

附录二引理 2 的证明

考虑一个 2-敏感度为 1 的查询 R。让 ***r 是添加高斯噪声量的机制*** $\mu := (1/[\sqrt{\max}$邻近数据集 五(丁、丁)])到 r。**根据推论 1，随机化机制 R 相对于相邻数据集 D 和 D 的隐私损失是由** $\mathcal{N}([U(D,D')/2], U$ **(D，D))为 U(D，D)** $:= ([\|R(D) - (D')\|_2]^2/\mu^2)$**。通过考虑 R 的 2-灵敏度(即，。*研发-研发)2)作为 1，最大相邻数据集 D，D 五(丁 、丁)和最大邻近数据集 D，D 大学(博士，博士)都一样。*** 此外， 从 定理5 关于 [65], 出租 Y (resp。

r)满足(，δ)-差分隐私可以转换为最大相邻数据集 D，DV(D，D)(分别)上的条件。 最大邻近数据集 大 学(博士，博士)). 那么，让 Y 满足 (ϵ, δ) -差异隐私就等于让 R 坐 - isfy(，δ)-差异隐私。从引理 1，R 达到 /-差分隐私，μ =高斯(1，，δ)；即。如果最大相邻数据集 五(丁、丁)=[高斯(1，，δ)]2.综上所述，我们完成了引理 2 的证明。

附录三定理 2 的证明

我们用定理 1 来说明定理 2 的结果① ②和③。①证明: 在 2A)和 C)的情况下，Qm 可以重用以前的噪声。因此，隐私损失仍然是)/2]，A(D，D))根据(5)。

②的证明:在情况 1)中，Qm 不能重用以前有噪声的答案，新噪声跟随 N(0，σm)。 因此，

乙(丁、丁):= A(D，D)) $+ ([\|Q_m(D) - Q(D')\ _2]^2/\sigma_m{}^2)$ ***m*** 。

③证明:在 2B 的情况下，Qm 可以重用以前的噪声 部分回答，所以我们可以用(5)证明。然后，引理 2 进一步暗示定理 2 的结果。

④证明:Qm 可以完全复用案例 2A)和 C)中的旧噪声结果。 因此，隐私级别不会改变。

⑤的证明:从引理 2，我们有最大邻数据集 sD，D 甲(丁，丁) = **[高斯(1，旧，δ 预算)]2 和 最大邻近数据集 {甲(丁，丁) + [Qm(D) − 质量管理)2]2 × (1/σm2)}**

=[高斯(1，新，δ 预算)]2。上述两个方程产生[高斯(1，新，δ 预算)]2

[高斯(1，旧，δ 预算)]2 =最大邻近数据集 D，D [Qm(D)-Qm(D)2]2 × (1/σm2) = ***qm2×(1/σm2)=σm2.*** **因此，高斯(Qm，_squared_cost，δbugdet) = σm。**

⑥的证明:从引理 2，我们得到了最大邻数据集 sD，D 甲(丁，丁)=[高斯(1，旧，δ 预算)]2 和

!

邻近数据集 D，D 甲(丁，丁

$$)+\left[\|Q_m(D)-Q_m(D')\|\ \right]^2{}_2 \times\left[\frac{1}{\sigma_m{}^2}-\frac{1}{[\min(\boldsymbol{\Sigma}_t)]^2}\right]\Bigg\}$$

$$=\text{高斯 1，新，δ 预算})\Big]^{-2}。$$

以上两个等式得出

高斯 1，新，δ 预算)]−2 − [高斯 1，旧，δ 预算 −2 $\left[\|Q_m(D)-Q_m(D')\|_2\right]^2$ **=最大相邻数据集**

$$\times\left[\frac{1}{\sigma_m{}^2}-\frac{1}{[\min(\boldsymbol{\Sigma}_t)]^2}\right]$$

$$=\text{Qm2}\times\left[\frac{1}{\sigma_m{}^2}-\frac{1}{[\min(\boldsymbol{\Sigma}_t)]^2}\right]。$$

然后利用引理 1 中高斯(Q，，δ)的表达式，进一步得到结果⑥。

附录四定理 3 的证明

首先，从定理 2 开始，在算法 1 被用于回答所有 n 个查询并且查询 Qi 在(I，δI)-差分隐私下被回答之后，相对于相邻数据集 D 和 D 的总隐私损失 由下式给出)/2]，G(D，D))对于一些 G(D，D).

接下来，我们用定理 2 进一步说明 G(D，D)由(6)给出。根据定理 2，在所有查询中，只有属于情形 1)和 2B)

的查询对 G(D，D)有贡献). 下面我们分别讨论贡献。N1 表示 I∞{ 1，2，...,。*n}这样齐在例 1)中，我们从定理 2 的结果②知道例 1)中的查询对 G(D，D)的贡献)由 下 式给出*

$$\sum_{i\in N_1}\frac{\left[\left\|Q_i(D)-Q_i(D')\right\|_2\right]^2}{\sigma_i^2}. \text{ (25)}$$

下面我们使用定理 2 的结果③来计算案例 2B 中的查询对 G(D，D)的贡献). 对于案例 2B 中的查询类型集 T2B，我们分别讨论了每个查询类型 t ∈ T2B。

从定理 2 的结果③来看，对 G(D，D)通过回答 Qjt，差分隐私下的 1 为

$$\left[\left\|Q_{j_{t,1}}(D)-Q_{j_{t,1}}(D')\right\|_2\right]^2\left(\frac{1}{\sigma_{j_{t,1}}^2}-\frac{1}{\sigma_{j_{t,0}}^2}\right).$$

同样，对 G(D，D)的贡献)通过回答 Qjt，差分隐私下的 2 是

$$\left[\left\|Q_{j_{t,2}}(D)-Q_{j_{t,2}}(D')\right\|_2\right]^2\left(\frac{1}{\sigma_{j_{t,2}}^2}-\frac{1}{\sigma_{j_{t,1}}^2}\right).$$

在案例 2B 中，对额外的类型 t 查询重复类似的分析)。特别地，对于每个 s∞{ 1，2，...,。*mt}，对 G 的贡献(D，D)通过回答 Qjt，差分隐私下的 s 为*

$$\left[\left\|Q_{j_{t,s}}(D)-\bullet\ \text{Qjt，s}\cdot\right.\left.(D')\right\|_2\right]^2\left(\frac{1}{\sigma_{j_{t,s}}^2}-\frac{1}{\sigma_{j_{t,s-1}}^2}\right)。\text{ (26)}$$

mt}，我们得到对于每个查询类型 t ∈ T2B，对 G(D，D)通过

差分隐私下的 Qjt、mt 为

$$\sum\ \left[\|_{\text{Qjt，s(D)Qjt，s}}\right.$$

mt} σjt，s σjt，S1

，，$Q_{jt,mt}$ for $j_{t,0}$,$j_{t,1}$,...jt 的 Qjt，mt，0，jt，1，...,。jt，mt 都是类型查询，Qjt，s(D)-Qjt，s(D)2 对于 s∑{ 1，2，...,因此，我们把(27)写成

$$\frac{}{(D)-Q_{j_{t,s}}(D')\|_2\right]^2}\sum\ \left\{\left[\|\right.$$

mt} σj2t，s

$$-\frac{\left[\left\|Q_{j_{t,s-1}}(D)-Q_{j_{t,s-1}}(D')\right\|_2\right]^2}{\sigma_{j_{t,s-1}}^2}\Bigg\}$$

$$\underline{\qquad\qquad\qquad\qquad}$$

$$\underline{\qquad\qquad\qquad\qquad}\text{山}\left\|(D')\right\|_2\right]^2$$

σ j2t，mt

$$\frac{\left[\left\|Q_{j_{t,0}}(D)-Q_{j_{t,0}}(D')\right\|_2\right]^2}{}. \text{ (28)}$$

− σj2t，0

t ∈ T2B 的所有(28)相加，对 G(D，D)通过回答案例 2B 中的所有问题)是

$$(D')\|_2\right]^2\text{山}$$

$$\sum\left\{\left[\|\right.\right.$$

T2B σj2t，mt

$$-\frac{\left[\left\|Q_{j_{t,0}}(D)-Q_{j_{t,0}}(D')\right\|_2\right]^2}{\sigma_{j_{t,0}}^2}\Bigg\}. \text{ (29)}$$

然后，G(D，D)作为(25)和(29)的和由(6)给出。

综上所述，我们已经证明了在使用算法 1 来回答差分隐私下的所有 n 个查询之后，相对于相邻数据集 D 和 D 的总隐私损失 由下式给出)/2]，G(D，D))为 G(D，D)在(6)中。此外，在

最大

$$\left\|Q_i(D)-Q_i(D')\right\| =\text{相邻数据集2 齐和}$$

$$(D')\|_2$$

最大 Qjt，mt(D)-Qjt，mt 相邻数据集 D，D

$$=_0(D)-Q_{jt\,0}(D')\| =/\text{相邻数据集}$$

max D，D Qjt， ， 2 (t 型)，

我们使用(6)得到最大邻近数据集 D，D G(D，D)由(7)给出。

最后，从引理 2，我们的总隐私成本算法 1 可以由(我们的，δ 预算)-差分隐私给出，以使我们满意

高斯 1，我们的，δ 预算)]$^{-2}$ = 最大总设计师，总设计师 相邻数据集

或满足任何和 δ 的(ϵ,δ)差分隐私

[高斯(1，，δ)]2 =最大邻近数据集 D，D G(D，D).

附录五高斯机制的效用

证据:一维查询 Qm 的噪声响应是 $\widetilde{Q}_m(D) =$ Qm(D) + N(0，σ2)。设 Qm(D)-Qm(D)p≤α 的概率为 1β，那么我们有

$$\begin{aligned}1-\beta &= \mathbb{P}\left[\left\|\widetilde{Q}_m(D)-Q(D)\right\|_p \le \alpha\right]\\ &= \mathbb{P}\left[\left|N\left(0,\sigma^2\right)\right| \le \alpha\right]\\ &= \mathbb{P}\left[-\alpha \le N\left(0,\sigma^2\right) \le \alpha\right]\\ &= \mathbb{P}\left[N\left(0,\sigma^2\right) \le \alpha\right]-\mathbb{P}\left[N\left(0,\sigma^2\right) \le -\alpha\right]\\ &= \frac{1}{2}\left[1+\operatorname{erf}\left(\frac{\alpha}{\sigma\sqrt{2}}\right)\right]-\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{-\alpha}{\sigma\sqrt{2}}\right)\right]\\ &= \operatorname{erf}\left(\frac{\alpha}{\sigma\sqrt{2}}\right),\end{aligned} \quad (30)$$

其中 erf()表示错误函数，而(30)的最后一步使用了 erf()是奇数函数的事实。

根据高斯分布的二西格马规则[66]，也可以从上面的等式中获得，95%的值位于平均值的两个标准偏差内。■ 因此，如果我们设置 α = 2σ，β ≈ 0.05。

承认

本材料中表达的任何观点、发现和结论或建议都是作者的观点、发现和结论或建议，不反映新加坡国家研究基金会的观点。

参考

[1] 动词 （verb 的缩写)。都铎王朝，v。Gulisano。阿尔格伦和 m。Papatriantafilou，“北京谱仪:大规模物联网系统的差异私有事件聚合”，未来一代。*电脑。系统。*卷。108 页。7 月 1241 日至 1257 日。2020.

[2] 米（meter 的缩写))。单位。m .哈桑。H. 雷马尼和 j。陈，物联网系统中的隐私保护:集成问题、前景、挑战和未来研究方向。*电脑。系统。*卷。97 页。8 月 512 日至 529 日。2019.

[3] J. 熊等，。“增强物联网智能电力服务中数据集群的隐私性和可用性”，IEEE 物联网杂志，。卷。6，没有。第 2 页。4 月 15 日 15:30–15:40。2019.

[4] J. 刘，c。张和 y。方，“EPIC:保护智能家居免受互联网流量影响的差异化隐私框架分析”，IEEE 物联网 j，。卷。5，没有。第 2 页。4 月 1206 日至 1217 日。2018.

[5] K. 盖，y。吴，l。朱，z。张和 m。邱，“基于工业物联网的差异化隐私保护”，IEEE 译。*Ind。信息。*卷。16，不。第 6 页。4156–4165，六月。2020.

[6] C. Dwork。Kenthapadi。麦克雪莉，我。米罗诺夫和 m。Naor，“我们的数据，我们自己:通过分布式噪声产生的隐私”，载于 Proc。*里面的。糖膏剂。理论应用。Cryptogr。Techn。(EUROCRYPT)，2006 年，第 10-11 页。*486–503.

[7] C. Dwork，f。McSherry。尼西姆和 a。史密斯，“校准私人数据分析中的噪声灵敏度”，载于《过程》。*密码理论。糖膏剂。(土耳其竞争委员会)，2006 年，第 10 页。*265–284.

[8] C. Dwork 和 a。罗斯，“差异隐私的算法基础”，发现。*趋势理论。电脑。Sci。*卷。9 个，编号。第 3-4 页。211–407, 2014.

[9] C. Dwork 和 g。名词（noun 的缩写)。Rothblum，“集中差别隐私”，2016 年.[在线]。可用:arXiv:1603.01887。

[10] 米（meter 的缩写))。小面包和 t。Steinke，“集中差分隐私:简化、扩展和下限”，在 Proc 中。*密码理论。糖膏剂。(土耳其竞争委员会)，2016 年，第 10 页。*635–658.

[11] C. Dwork，v .费尔德曼，m。哈特，t。皮塔西，o。莱因戈尔德和 a。罗斯，“自适应数据分析和保持重用中的一般化”，载于《过程》。*糖膏剂。神经 Inf。流程。系统。(国家实施计划)，2015 年，第 100 页。*2341–2349.

[12] F. 麦克雪莉和 k。塔尔瓦尔，“通过差别隐私的机制设计”，在 Proc。*IEEE Symp。找到了。电脑。Sci。(FOCS)，2007 年，第 10 页。*94–103.

[13] 米（meter 的缩写))。Abadi 等人，。Proc 中的“具有差异隐私的深度学习”。*ACM Conf。电脑。社区。《安全》(CCS)，2016 年，第 10 页。*308–318.

[14] J. 唐，a。科罗洛娃，x。白，x。王，还有 x。王，“苹果在 MacOS 10.12 上实施差异化隐私的隐私损失”，2017.[在线]。可用:arXiv:1709.02753。

[15] Ú. v .埃尔林松。宾虚和甲。科罗洛娃，“RAPPOR:随机化的可聚合的隐私保护序数反应”，载于 Proc。*ACM Conf。电脑。社区。《安全》(CCS)，2014 年，第 10-11 页。*1054–1067.

[16] B. 丁，j。库尔卡尼和 s。叶卡宁，“私下收集遥测数据”，在 Proc。*糖膏剂。神经 Inf。流程。系统。(国家实施计划)，2017 年，第 10 页。*3571–3580.

[17] R. 亨利，a。赫尔茨贝格和 a。凯特，“区块链访问隐私:挑战和方向”，IEEE 安全隐私，第一卷。16，不。第 4 页。7 月 38 日至 45 日。/8 月。2018.

[18] J. 康，z。熊，d。Niyato，s .谢，还有 j。张，“可靠联合学习的激励机制:一种结合声誉和契约理论的联合优化方法”，IEEE 物联网 J，。卷。6，没有。第 6 页。10700–10714，12 月。2019.

[19] 米（meter 的缩写))。杨，a。Margheri。胡，与伏。萨松，“与区块链的云联盟中不同的私有数据共享”，IEEE 云计算。卷。5，没有。第 6 页。11 月 69 日至 79 日。/Dec . 2018.

[20] 南。中本聪。(2008). *比特币:一种点对点的电子现金系统。*[在线]。可用:https://bitcoin.org/bitcoin.pdf

[21] 投资媒体。*区块链。*[在线]。可查阅:https://www.investopedia。com/terms/b/blockchain.asp

[22] G. 伍德，“以太网:一个安全的分散的通用交易分类账”，瑞士祖格，以太网，黄皮书，2014 年。

[23] 名词（noun 的缩写)。萨伯。(1994). *智能合同。*[在线]。可查阅:http://www . fon . hum . UVA . nl/rob/Courses/informationinspech/CDROM/Literature/LotWinterschool 2006/Szabo . best . vwh . net/smart . contracts . html

[24] 动词 （verb 的缩写)。Buterin，“下一代智能合约和分散应用平台”，瑞士 Zug，Ethereum，白皮书，2014 年。

[25] 长度。米（meter 的缩写))。韩，y。赵，和 j。赵，“基于区块链的差异化隐私成本管理系统”，2020.[在线]。可用:arXiv:2006.04693。

[26] 页（page 的缩写)。凯鲁兹，s。哦，还有 p。维斯瓦纳特，“差分隐私的合成定理”，IEEE 翻译。*Inf。理论，卷。*63 号没有。第 6 页。4037–4049，六月。2017.

[27] *Ropsten 的官方 Github 页面。*访问时间:1 月。9, 2019.[在线]。可用:https://github.com/ethereum/ropsten

[28] D. 桑切斯，j . 多明戈-费雷尔和 s。马丁内斯，“通过单变量微聚集改善差别隐私的效用”，载于《程序》。*里面的。糖膏剂。隐私统计。数据库，2014 年，页。*130–142.

[29] 名词（noun 的缩写)。王等。“收集和分析具有局部差异隐私的多维数据”，载于 Proc。*IEEE 第 35 届国际。糖膏剂。数据工程。(ICDE)，2019 年，第 10 页。*638–649.

[30] *加纳切。*查阅时间:2020 年.[在线]。提供:https://www.trufflesuite。com/ganache

[31] A. 热尔韦，g。k .卡拉姆。Wüst，v。Glykantzis，h . Ritzdorf 和 s。凯昆，“关于工作证明的安全性和性能，区块链”，在 Proc。*ACM*

SIGSAC Conf。电脑。社区。《安全》(CCS)，2016 年，第 10 页。 3–16.

[32] B. 曹等。“当物联网遇到区块链:分布式共识的挑战”，IEEE Netw。卷。33 号没有。第 6 页。11 月 133 日至 139 日。/Dec . 2019.

[33] D. 华盛顿特区。Jagodi，c 和 s。区块链技术，比特硬币和以太网:一个简单的概述。*第 17 个 Int。辛普。Infoteh-Jahorina (Infoteh)，2018 年，第 10 页。* 1–6.

[34] J. Stark，“理解 Ethereum 的第 2 层缩放解决方案:国家通道、等离子体和 truebit”，2018.[在线]。可用:https://medium . com/l4-media/making-sense-ether eums-layer-2 scaling-solutions-state-channels-plasma-and-true bit-22cb 40 dc2 f 4

[35] H. 唐，y。史，和 p。董，“基于熵和 TOPSIS 的公共评价”，专家系统。*应用，。*卷。第 117 页。3 月 204 日至 210 日。2019.

[36] 米（meter 的缩写））。H. 米拉兹和 d。C. 唐纳德，“LApps:区块链闪电网络应用的技术、法律和市场潜力”，载于 Proc。*第三个 Int。糖膏剂。Inf。系统。数据最小值。* 2019 年，共页。185–189.

[37] A. Blum。利格特和 a。非交互式数据库隐私的学习理论方法。*ACM，第一卷。* 60，不。第 2 页。1–25, 2013.[38] Apache JMeter。访问时间:1 月。20, 2021.[在线]。可用:https://jmeter.apache.org/

[39] 页（page 的缩写）。郑，z。郑，x。罗，x。陈，和 x。刘，系统的详细和实时性能监控框架。*IEEE/ACM 40 Int。糖膏剂。软。英格。软。英格。普拉特。Track (ICSE-SEIP)，2018 年，第 100-100 页。* 134–143.

[40] 南。森岛和 h。松谷，“使用图形处理器加速全节点的区块链搜索”，载于 Proc。*第 26 届欧微国际。糖膏剂。并行发行版。Netw。基于流程。(PDP)，2018 年，页。* 244–248.

[41] *活动监视器用户指南。*访问时间:2 月。2, 2021.[在线]。可用:https://support.apple.com/en-sg/guide/activity-monitor/欢迎/mac

[42] 南。国王和 s。纳达尔，“PPCoin:点对点加密货币与证据”，自出版论文，第一卷。第 19 页。8 月 1 日。2012.

[43] X. 李。李。颜，z。程，w。孙和 h。朱，“利用高维高斯机制缓解回归模型的查询泛滥参数复制攻击”，2020.[在线]。可用:arXiv:2002.02061。

[44] J. 康，z。熊，d。Niyato。叶，马超。爱达荷（Idaho 的缩写）。金和杰。赵，“走向安全的支持的车辆互联网:使用声誉和契约理论优化共识管理”，IEEE 翻译。*Veh。技术。*卷。68，不。第 3 页。3 月 2906 日至 2920 日。2019.

[45] J. 康等。“车载边缘计算和网络中安全高效数据共享的区块链”，IEEE 物联网杂志，。卷。6，没有。第 3 页。4660–4670，六月。2019.

[46] Y. C. 蔡英文。曹先生，z。-是的。刘和 k。陈，“分散应用的改进的非交互式零知识范围证明”，载于 Proc。*IEEE Int。糖膏剂。分散应用。基础设施。(DAPPCON)，2019 年，第 10 页。* 129–134.

[47] A. c .费尔南德斯·安塔。乔治乌和纽约。尼科劳，“原子追加:销售汽车和用多个分布式分类账协调军队”，2019 年。

[48] 问。冯，d。他，s。玉米，m。K. 可汗，还有 n。库马尔，“区块链系统隐私保护调查”，j。*Netw。电脑。应用，。*卷。126 页。1 月 45 日至 58 日。2019.

[49] G. Zyskind，o . 内森和 a。南。彭特兰，“分散隐私:利用区块链保护个人数据”，载于《程序》。*IEEE 安全隐私研讨会(SPW)，2015，页。* 180–184.

[50] 问。夏，鄂。B. Sifah，a。斯玛希，s。阿莫法，和 x。张，:“云环境下基于区块链的电子病历数据共享”，《信息》，第 1 卷。8，没有。第 2 页。44, 2017.

[51] Y. 鲁问。唐，和 g。王，“ZebraLancer:开放区块链上的私有匿名众包系统”，载于 Proc。*IEEE 第 38th Int。糖膏剂。发行版。电脑。系统。(国际发展合作中心)，2018 年，第 10-11 页。* 853–865.

[52] 南。胡，c。蔡，问。王，c。王，x。罗，和 k。任，“搜索加密云与相遇:一种分散、可靠和公平的实现”，载于 Proc。*IEEE INFOCOM Conf。电脑。社区。* 2018 年，共页。792–800.

[53] 米（meter 的缩写））。洛昂哥和 c。Pon，“保留网络:公共区块链的隐私层”，保留网络。代表，。2018.[在线]。可用:https://keep。网络/白皮书

[54] J. 埃雷拉-琼科马尔蒂和 c。佩雷斯-索拉，“比特币交易中的隐私:来自区块链可扩展性解决方案的新挑战”，载于《人工智能建模决策》。瑞士湛:施普林格，2016 年，页。26–44.

[55] A. 南。Sani 等人，。“Xyreum:用于 IIoT 安全和隐私的高性能和可扩展的区块链”，在 Proc 中。*IEEE 第 39 届国际。糖膏剂。发行版。电脑。系统。(国际发展合作中心)，2019 年，第 10-11 页。* 1920–1930.

[56] X. 小，g。m .本德。干草和 j。盖尔克，“信息产品:减少相对误差的差异化隐私”，载于《过程》。*ACM Int。糖膏剂。管理。数据(SIGMOD)，2011 年，第 10 页。* 229–240.

[57] C. 李和 g。Miklau，“一种在差异隐私下准确回答查询的自适应机制”，Proc。*VLDB 捐赠基金，第一卷。* 5，没有。第 6 页。514–525, 2012.

[58] G. Kellaris 和 s。帕帕多普洛斯，“通过分组和平滑的实用差别隐私”，Proc。*VLDB 捐赠基金，第一卷。* 6，没有。第 5 页。301–312, 2013.

[59] G. 雅罗斯拉夫采夫，g。Cormode，c。米（meter 的缩写））。Procopiuc，和 d。Srivastava，“数据立方体和应急表的精确和有效的私有发布”，在 Proc 中。*里面的。糖膏剂。数据工程。(ICDE)，2013 年，第 10 页。* 745–756.

[60] F. 麦克雪莉和 r。Mahajan，“差分专用网络跟踪分析”，ACM SIGCOMM Comput。*社区。Rev，。*卷。40，不。第 4 页。123–134, 2010.

[61] 页（page 的缩写）。Aaby。米（meter 的缩写））。De Acuna。麦克法兰和 w。J. 布坎南，“恶意软件数据集上使用 RAPPOR 的隐私参数变化”，载于 Proc。*第 17 届 IEEE 国际会议。糖膏剂。信任安全隐私计算。社区。第 12 届 IEEE 国际。糖膏剂。大数据科学。英格。(TrustCom/BigDataSE)，2018 年，第 10-11 页。* 938–945.

[62] A. 科斯巴，a。米勒，e。史，z。文，和丙。帕帕曼图，“霍克:密码学和隐私保护智能合同的区块链模型”，载于《程序》。*IEEE Symp。安全隐私(SP)，2016 年，共页。* 839–858.63 权威证明。查阅时间:2020 年.[在线]。可用:https://github。com/paritytech/parity/wiki/权威证明-链

[64] A. Agarwal。赫利希，s。卡马拉和 t。Moataz，“用于差异隐私的加密数据库”，Proc。*隐私增强。技术。*卷。2019 年没有。第 3 页。170–190, 2019.

[65] B. Balle 和 y。-X。王，“改进差分隐私的高斯机制:分析校准和最佳去噪”，载于 Proc。*里面的。糖膏剂。马赫。学习。(ICML)，2018 年，第 10 页。* 394–403.

[66] F. 普凯尔西姆，“三西格玛法则”，阿米尔。*统计一下。*卷。48，不。第 2 页。88–91, 1994.

杨钊(IEEE 研究生会员)于 2015 年获得新加坡国立大学电气工程硕士学位。他目前正在攻读博士学位。新加坡南洋理工大学计算机科学与工程学院学位。

她的研究兴趣包括联合学习、区块链、差分隐私和 6G。

赵骏(IEEE 成员)于 2010 年获得中国上海交通大学学士学位和博士学位。美国宾夕法尼亚州匹兹堡卡内基梅隆大学电气和计算机工程学位。格利戈。亚甘；合作者:a。佩里格)，2015 年加入 CMU 著名的共青团安全与隐私研究所。

在加入之前，他先是和肖一起做博士后，然后是教职员工，他是美国亚利桑那州坦佩市亚利桑那州立大学的博士后，是亚利桑那计算博士后最佳实践研究员(顾问:j。张，v。可怜的)。现任新加坡南洋理工大学计算机科学与工程学院助理教授。他的研究兴趣包括通信、网络、安全和人工智能。

康获得硕士和博士学位。分别于 2015 年和 2018 年获得中国广州广东工业大学学位。现为新加坡南洋理工大学博士后。他的研究兴趣主要集中在区块链、无线通信中的安全和隐私保护以及网络。

张泽航于 2019 年获得中国广州广东工业大学硕士学位。

他目前是新加坡南洋理工大学的研究助理。他的研究兴趣主要集中在区块链、互联网数据安全和隐私保护。

杜斯特·尼耶托(美国电气和电子工程师协会研究员)获得了学士学位。1999 年获得泰国曼谷国王蒙古技术学院的学位和博士学位。2008 年获得加拿大不列颠哥伦比亚省温尼伯市马尼托巴大学电气和计算机工程学位。

现任新加坡南洋理工大学计算机科学与工程学院教授。他的研究兴趣是无线通信、物联网和能源收集领域

传感器网络。

施淑玉(IEEE 成员)于 2011 年获得中国合肥中国科学技术大学学士学位，并获得博士学位。2011 年获得日本 Hayama SOKENDAI 学位。

她目前是中国南京南京大学计算机科学系的研究副教授。2016 年至 2018 年任新加坡南洋理工大学计算机科学与工程学院并行与分布式计算中心无线与网络化分布式传感系统组研究员.她在国家信息学研究所和索肯岱的信息学系工作。2015 年 4 月至 2016 年 10 月，她还是 JSPS 研究员.她的研究兴趣集中在移动和无处不在的计算上。

林国彦(电气及电子工程师学会资深会员)获学士学位。学位(一级荣誉)。1987 年获英国伦敦大学计算机科学博士学位。1990 年获得英国剑桥大学学位。

2002 年至 2010 年，他是中国北京清华大学的教授.自 1990 年以来，他一直是新加坡国立大学和伦敦大学的教员。现任新加坡南洋理工大学计算机科学与工程学院教授。他是剑桥大学艾萨克·牛顿研究所的客座科学家，也是美国 DC 华府欧洲研究所的系统安全客座教授。他的研究兴趣包括分布式系统、物联网安全基础设施、区块链分布式协议、生物特征加密、国土安全和网络安全。

教授。林氏于一九九八年获日本工商会颁发新加坡基金奖，以表扬他在新加坡资讯保安的研究及发展成就。