This paper appears in ACM ASIA Conference on Computer and Communications Security (ACM ASIACCS) 2020. Please feel free to contact us for questions or remarks.

2020 年 10 月 5 日至 9 日在台湾台北举行的第 20 届亚洲气候大会

基于区块链的差异化隐私成本管理系统

莱昂·韩梅* 杨钊* 赵骏
南洋理工大学南洋理工大学 南洋理工大学
新加坡新加坡 MLEONG008 @ e . NTU . edu . SG s 180049 @ e . NTU . edu . SG 赵骏@ntu.edu.sg。

摘要

隐私保护是各个部门非常关心的问题。为了保护个人用户数据，一项新兴技术是差别隐私。然而，对于频繁查询的数据集，它仍然存在局限性，如隐私成本积累过快。为了解决这个限制，本文探讨了一个安全的分散分类账的整合，区块链。区块链将能够跟踪使用差分隐私算法生成的所有有噪声的响应，并允许某些查询重用旧的响应。在本文中，一个基于区块链的隐私管理系统的演示被设计为一个交互式的分散式网络应用程序(DApp)。创建的演示说明了利用区块链将允许累积的总隐私成本显著降低。

CCS CONCEPTS

安全性和隐私→分布式系统安全性；隐私保护协议；计算机系统组织→对等体系结构。

关键词

差别隐私，区块链。

自动呼叫管理参考格式:
韩梅、杨钊和赵骏。2020.基于区块链的差异化隐私成本管理系统。在2020年10月5日至9日举行的第十五届亚洲计算机与通信安全会议上，*台湾台北*。美国纽约州纽约市 ACM，3 页。https://doi.org/10.1145/1122445.1122456

1 介绍

在今天的数字时代，互联网给了我们收集和获取各种信息的潜力。个人愿意为了网上方便而泄露个人信息[1]。关于个人的大量数据不断存储在各种数据库中。当数据得到有效分析时，它可以转化为关于个人行为的有意义的信息。这导致个人数据成为数字时代，受到每个行业和政府的追捧。尽管数据分析可以带来巨大的好处，但对敏感数据的不当处理通常会导致比预期披露更多的信息。保护用户隐私的一种尝试是发布匿名或聚合数据。然而，事实证明，当与其他数据源结合时，记录可以去匿名化。当攻击者能够准确地将匿名数据库与另一个非匿名数据库匹配时，就会发生这种情况[6]。

差异隐私的引入使我们更接近于实现保护个人隐私的目标，同时仍然揭示关于数据集的有意义的信息。差分隐私背后的简单想法是在输出结果中加入一些噪声，以便在数据集中增加或不增加单个输入时，噪声不会发生显著变化。(ϵ,δ)-Differential 隐私[3]定义为:

定义 1。*随机算法满足(ϵ,δ)-Differential 隐私 if，对于任何两个相邻数据集，相差一条记录 (d 和 d′)以及输出范围的所有子集 s，它生成随机化输出，使得:pr[y(d) ∈s] ≤eϵpr[y(d′) ∈s]+δ，其中概率空间在随机化算法 y 的硬币翻转上。*

传统上，用不同隐私处理的每个查询都会产生隐私成本。即使是相同的查询，这种隐私成本也会累积。它带来了隐私成本可能超过隐私预算的问题，导致更大比例的隐私泄露[2，4]。

区块链技术的兴起为上述隐私预算耗尽的问题提供了一个可能的解决方案。通过利用区块链的分散、防篡改和可追溯的关键优势，它提供了一个分布式、可信的对等网络平台，以存储关于以不同隐私处理的查询的信息[7]。这些交易数据随后可以被检索以进行处理，以便可能重新使用先前生成的有噪声的答案。在随后的网络 DApp 演示中，将重点介绍旧的嘈杂答案的重用。使用基于区块链的隐私管理系统，总的隐私成本将大大降低，满足频繁查询的数据集，如病历数据集。

2 相关工程

Zyskind 等人。[9]解决使用第三方移动平台时的隐私问题，建议将区块链与区块链以外的存储结合起来，创建一个个人数据管理平台，以增强隐私。这允许用户拥有对其数据的所有权和控制权，而无需信任任何第三方。科斯巴等人。[5]提出了构建隐私保护智能合约的框架。霍克提出的框架，

---

*两位作者对论文的贡献相等。名字是按字母顺序排列的。

对应作者

2020 年 10 月 5 日至 9 日在台湾台北举行的第 20 届亚洲气候大会

2020 计算机械协会。美国计算机学会 ISBN 978-1-4503-XXXX-X/18/06… 15.00 美元
https://doi.org/10.1145/1122445.1122456

区块链和用户之间的加密协议。该加密协议包括使用经过验证的数据结构和零知识证明来增加安全性。

上述研究主要关注区块链和用户之间的身份隐私，同时信任区块链的匿名性。然而，缺乏文献不仅关注区块链的隐私，而且关注数据库本身的隐私保护。以前也做过重复使用有噪声的答案来保护隐私的研究。

肖等。[8]提出了一种差分私有算法，该算法将添加到不同查询结果中的拉普拉斯噪声相关联，以提高数据效用。然而，在本文中，高斯噪声被用于拉普拉斯噪声之上，以便更容易地对多种查询类型进行隐私分析。独立拉普拉斯随机变量的和不遵循拉普拉斯分布。另一方面，独立高斯随机变量的和仍然遵循高斯分布。因此，使用高斯噪声将允许算法处理不同类型的查询。

3 演示概述

为了说明所提出的基于区块链的隐私管理系统，创建了一个分散的 web 应用程序。该演示模拟了系统如何在跟踪和降低隐私成本的同时实现差分隐私算法。图。图 1 显示了基于区块链的演示的整体用户界面。

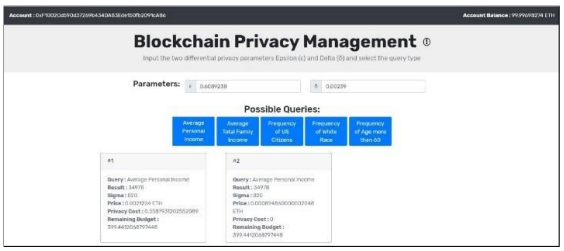


图 1:基于区块链的隐私管理系统演示截图。

3.1 Ethereum 帐户信息

图 2:显示用户的以太网帐户信息。

一旦用户授予演示权限，用户的 Ethereum 帐户信息就会显示在页面顶部固定的栏上。演示从浏览器的元掩码扩展中获取账户信息(钱包地址和里面的余额)。当连接到元掩码时，将出现一个由元掩码管理的弹出窗口，询问用户是否允许访问该帐户。只有在授予权限后，才会显示帐户信息。

图 3:获取用户期望的差异隐私参数。

3.2 定义不同的隐私参数

在演示中，用户将能够使用图 1 所示的输入栏指定用于差分隐私算法(ϵ 和 δ)的参数。3.ϵ 值用于确定隐私级别的严格程度。ϵ 值越小，隐私保护越好。 *δ 定义了 ϵ-differential 隐私概念的放松程度。*

3.3 查询选择

图 4:支持的查询类型。

用户可以通过按下创建的按钮来选择他们想要查询的查询类型。在后端，每个按钮都链接到一个预先计算好的敏感度级别，并根据其查询类型分配给该按钮。差分隐私算法根据查询函数的灵敏度来缩放生成的噪声。此敏感度级别是任何两个相邻数据集的真实查询结果之间的最大距离，两者相差一条记录。灵敏度等级的计算如下:

′ ′

$$\Sigma Q = \max_{\text{neiдhborиндD}} D, \; D \Sigma Q(D) Q(D) \Sigma 2,$$

其中 Q 是查询 Q 的敏感度级别，D 和 D 是相差一条记录的相邻数据集。

3.4 重复使用先前的高斯噪声

该算法被设计为在标准偏差比较的基础上运行，并遵循图 1 中的一般工作流程。5.

基于来自用户的隐私参数，将计算 σ 值。这是该查询所需噪声的标准偏差。要回答灵敏度为 Qm 的查询 Qm，零-

将标准偏差 $\sigma = 2\ln$ δm ϵm 的 q 1.25×ΣQM 平均高斯噪声添加到真实查询结果中。由于高斯噪声可以从标准偏差中计算出来，因此所提出的系统存储标准偏差，并将其用作重复使用噪声的比较基础。系统从区块链检索所有以前的交易并进行比较。它首先检查区块链是否有任何使用相同查询类型和标准偏差的现有记录。如果在区块链找到现有记录，算法将返回与差分隐私输出相同的结果。如果没有找到现有记录，它将比较新查询和以前查询的标准偏差。该算法通过将噪声注入到先前的噪声中来重新使用高斯噪声，以生成满足隐私要求的新噪声。因此，如果新的标准偏差是

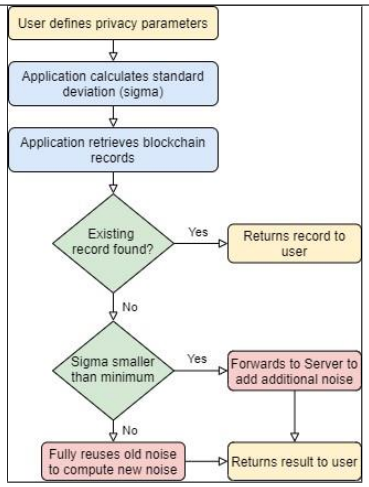
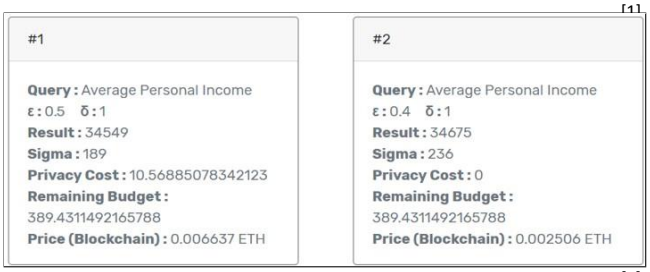


图 5:噪声重用的工作流程。

大于来自同一类型的所有先前查询的最小标准偏差。然后，该算法将计算需要添加到先前结果中的新噪声，并得出新结果。如果新的标准偏差小于先前生成的最小标准偏差，它将无法完全重用任何先前的噪声。该算法只能重用先前噪声的一小部分，并计算要添加的附加噪声。然后，查询将被转发到服务器，以将计算出的噪声添加到部分重用的噪声响应中。



3.5 输出显示

图 6:显示带有 ϵ 隐私成本的输出。

该演示显示了以卡片格式执行的任何查询的输出，如图 1 所示。6.一旦输出可用，这些卡片就会弹出。卡片的标题包含为提交的查询生成的查询标识。正文中包含请求的查询类型、应用程序生成的嘈杂响应、计算的标准差(σ)、区块链价格、隐私成本和剩余隐私预算。显示的隐私成本是 ϵ 隐私成本。

4 履行

该系统是利用 Bootstrap、Ethereum、MetaMask、Web3.js、Truffle Suite、Provable、MongoDB 开发的和英雄库。图。图 7 显示了建议系统的架构图。

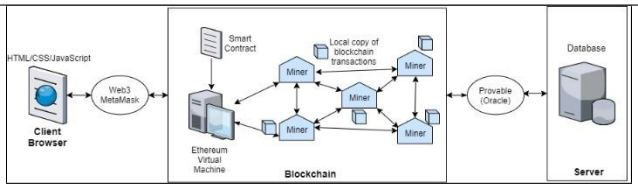


图 7:整体软件架构。

DApp 由运行在分散平台上的前端客户端浏览器和后端组成，在托管服务器上有 Ethereum 和 MongoDB 数据库。块菌套件和可证明用于开发智能合同。对于前端客户端浏览器，index.html 和 app.css 文件为用户定义了网页显示。它还利用 Bootstrap 与用户进行响应性交互。前端还包含一个 app.js 文件，该文件映射来自 index.html 的项目，与 Web3.js 交互，处理计算并将其解析到区块链。

5 结论

在本文中，开发了一个简单的 DApp 演示来说明使用区块链来重用以前从差分隐私算法生成的噪声响应。将来，这个演示可以通过添加图形元素来改进，以更好地显示系统的效果，例如用户可以提交的查询数量增加，而不会超过隐私预算。该演示还可以通过分析和量化隐私保护的程度以及噪声的重用来进一步改进。

参考

[1] 特伦斯·克雷格和玛丽·勒德洛夫。2011.隐私和大数据:参与者、监管者和利益相关者。“奥莱利媒体公司。".
弗雷德里克·库彭斯、法特玛·布瓦特尔和费滕·本·弗雷德里克。2019.差别隐私中隐私预算的最优分配。《互联网和系统的风险与安全:第 13 届国际会议，危机 2018》，法国阿卡川，2018 年 10 月 16-18 日，修订论文选集，第一卷。11391. 斯普林格 222 号。
辛西娅·德沃克、弗兰克·麦克谢里、科比·尼西姆和亚当·史密斯。2006.校准私人数据分析中的噪声灵敏度。密码学理论会议。斯普林格，265–284。
贾静瑜、、岳过、高继强、、彭然、。2019.具有预算选项的数据库查询系统，用于防止重复攻击的差异隐私。新计算环境中的安全和隐私国际会议。斯普林格，46-57 岁。
[5] 艾哈迈德·科斯巴、安德鲁·米勒、伊莲·施、子恺文和查拉姆波斯·帕帕曼图。2016.霍克:密码学和隐私保护智能合同的区块链模型。在 2016 年 IEEE 安全和隐私研讨会上。IEEE，839–858。
[6] 阿尔温德·纳拉亚南和维塔利·什马蒂科夫。2008.大型稀疏数据集的鲁棒去匿名化。2008 年 IEEE 安全与隐私研讨会。IEEE，111–125。
[7] 卡里姆·苏丹、乌马尔·鲁伊和鲁比娜·拉哈尼。2018.概念化的区块链:特征和应用。*arXiv 预印本 arXiv:1806.03693 (2018)*。
[8] 肖、本德、海和盖尔克。2011.信息产品:减少相对误差的差异化隐私。在 2011 年 ACM SIGMOD 国际数据管理会议记录中。229–240.
[9] 盖伊·齐斯金、奥兹·内森等人。2015.分散隐私:利用区块链保护个人数据。2015 年 IEEE 安全和隐私研讨会。IEEE，180–184。