双服务器模型中的分布式、私有、稀疏的直方图

詹姆斯贝尔
谷歌
英国伦敦
jhbell@google.com

阿德里亚加
谷歌
纽约，纽约，美国
adriag@google.com

巴迪加齐
谷歌
山景城，加州，美国
巴迪。
ghazi@gmail.com

拉维库马尔
谷歌
山景城，加州，美国
拉维。k53@gmail.com

帕辛马努朗斯
谷歌
泰国曼谷
pasin@google.com

玛丽安娜雷科娃
谷歌
纽约，纽约，美国
marianar@google.com

菲利普肖普曼
谷歌
纽约，纽约，美国
schoppmann@google.com

摘要

我们考虑了在安全多方计算（MPC）的双服务器模型中稀疏的，（e，6）-微分私有（DP）直方图的计算，该模型最近在聚合用户数据的隐私保护度量中获得了关注。

我们引入了一些协议，使两个半诚实的非合并服务器能够计算多个用户持有的数据上的直方图，同时只学习数据的私有视图。我们的解决方案实现同样的渐近ℓ_∞-误差为$O\left(\frac{\log(1/\delta)}{\varepsilon}\right)$如在中心模型中的DP，但不依赖于一个值得信赖的馆长。我们的协议的服务器通信和计算成本与直方图桶的数量无关，且与用户数量呈线性关系，而客户端成本与用户数量e和6无关。

它对用户数量的线性依赖使我们的协议能够很好地扩展，我们使用微基准测试确认了这一点：对于10亿用户，e = 0.5，和6=10-11.08，我们的协议的每个用户成本只有1 ms的服务器计算和339字节的通信。相比之下，使用混乱电路的基线协议只允许高达106用户，其中每个用户需要600 KB的通信。

中国化学会概念

安全和隐私的→隐私保护协议。

关键字

微分隐私，多方计算，直方图

ACM参考格式：

詹姆斯贝尔，阿德里亚加斯孔，巴迪加齐，拉维库马尔，帕辛马努兰西，玛丽亚娜雷科娃，和菲利普肖普曼。2022.双服务器模型中的分布式的、私有的、稀疏的直方图。在2022年ACM SIGSAC计算机和通信安全会议会议记录中(中国化学会

2022年11月7日至11日，美国加州洛杉矶。ACM，纽约，纽约，美国，22页。https://doi.org/10.1145/3548606.3559383

中国化学会‘22,2022年11月7日-11日，美国加州洛杉矶

ACM ISBN 978-1-4503-9450-5/22/11。
https://doi.org/10.1145/3548606.3559383

1介绍

在大用户群体中计算的汇总统计数据被广泛用于发现用户行为和偏好的一般趋势。应用程序可以在许多不同的上下文中找到，包括产品分析和浏览器遥测[10,15,26,27]，了解病毒[4,58]的传播，以及检测分布式攻击和欺诈行为[13,63]。在保护个人用户隐私的同时，设计高精度计算此类分析的技术一直是一个活跃的研究课题[6,10,11,13,15，
23，26，27，33，40，43，63，67–69，74，79]。

差异隐私（DP）[31,32]的概念形式化了一个算法的输出不会揭示关于个人用户贡献的实质性信息的保证。在计算过程中注入噪声，这也会影响输出的精度。中央DP机制[32]提供了隐私保证和准确性之间最著名的权衡。然而，它们依赖于一个强有力的假设，即存在一个可信的管理员，它可以访问整个数据集。本地DP设置[32,35,55]通过将隐私机制分配给客户端，减轻了中央管理员的隐私影响，但这在精度[8,21]上的成本很高。

安全多方计算（MPC）[45,46,60,77]提供了一种技术，允许两个或多方联合计算一个依赖于他们的私有输入的函数，同时在计算过程中除了函数输出之外没有透露任何东西。在分布式设置中实现强隐私性和高精度的一个自然想法是使用MPC来执行中央DP机制[31]。然而，将此想法直接应用于计算聚合用户统计数据将需要在数据包含在聚合统计数据中的所有用户的设备上执行多轮协议。考虑到现有的大规模MPC实现[3,56,76]的高计算和通信开销，以及客户端设备不可预测的可用性模式，这种方法在数亿或数十亿美元的用户群中变得具有挑战性。

一个中间信任模型是外包的MPC模型，它避免了一个中心聚合器和完全分布式MPC的可伸缩性挑战。这里是聚合器的功能

被分成了少数不相勾结的政党。它们接收从客户端获得的秘密共享（或加密的）输入，然后使用它们之间的MPC协议计算所需的聚合统计信息。只要至少有一方保持诚实，客户的投入就会是私人的，只显示期望的总额。除了较低的通信和计算开销外，外包的MPC模型还可以处理客户端退出，因为通常只需要来自每个客户端的一条消息。两个计算服务器的特殊情况称为双服务器模型，它已经在[4,27]中应用于许多大型MPC部署。虽然拥有更多政党的诚实多数协议可以带来更好的效率，但确保诚实多数的假设确实成立仍然具有挑战性。另一方面，与两党对手相比，超过双方的不诚实多数MPC协议的MPC协议存在性能缺陷(见，e。g.，用于比较)。因此，在这项工作中，我们关注这个设置，包括两个非合并服务器和大量客户端，每个客户端只发送一个消息。

稀疏直方图。许多流行的聚合函数都可以用用户数据上的直方图来描述。在这里，每个用户都有一个来自域D的值，其目标是计算持有每个可能的输入值的用户数量。在许多设置中，用户贡献的域D远远大于输入中唯一值的实际数量，并且在某些设置中，它也大于用户总数。因此，所得到的直方图通常是稀疏的，i。e.，域中的大多数值的计数将为零。例如，包括在用户[16,57]持有的字符串中计算重击球者，在位置数据[25]中寻找通勤模式，或空间分解[25]。

在稀疏直方图的情况下，计算效率的问题变得更加明显——理想情况下，协议应该实现计算和通信的复杂性，独立于域大小|D|，只依赖于需要处理的贡献的数量。在寻找这样的协议时，要回答的第一个问题是，是否有一个中央DP机制的输出长度和计算成本独立于|D |。虽然向直方图中每一个可能的条目添加DP噪声的机制并不满足这一特性，但Korolova等人的工作。[57]提供了这样一个解决方案，它保证总是（隐式地将零计数）报告为零，并且只报告非空直方图位置的一个子集。

利用现有的MPC技术来实现[57]的中心DP机制，也带来了一系列的挑战。显然，需要客户端发送与|D | [26]成比例的输入的技术是不可取的。分布式点函数[17]将客户端计算和通信压缩到O（log |D |），可以作为频率谕发现稀疏直方图[16]中的非零位置。然而，这种方法会导致DP也是O（log |D |）的错误，比[57]更糟糕。据我们所知，没有一种有效的DP协议来计算稀疏直方图，以实现独立于|D |的错误，而不依赖于可信的管理员。

我们的贡献。在这项工作中，我们提出了在双服务器模型中计算稀疏直方图的分布式协议。我们的协议需要从客户端进行一次O（log（|D |））位的通信，并且两个服务器之间的通信是线性的在来自客户的捐款数量上。它提供（e，6)-DP为输出，$\ell\infty$-误差为$O\left(\frac{\log(1/\delta)}{\varepsilon}\right)$，匹配

在中心DP模型中可能的最佳边界。

我们的协议保证输出是DP；此外，他们还保证了每个服务器的视图满足一个DP的计算版本，称为SIM+ -CDP [65]。然而，与以前关于分布式DP协议的工作不同，我们明确地指定了在协议执行期间显示的DP泄漏。这使得可以对DP保证之外的不同方法进行比较，特别是允许区分纯MPC解决方案和揭示额外信息的协议。

我们的结果总结在以下非正式定理中。

定理1。考虑n个客户端，每个客户端都持有一对内迪∈D，vali∈[Δ]。有一个回旋协议依赖于两个非合并服务器P1，P2为P1获得输入数据的直方图与l∞-error O（Δ log（1/6）/）。ε协议的输出直方图及其泄漏的组合（正式定义请参见定义3）为（e，6) -DP。对于n中的常数e和6逆多项式，服务器的通信和计算为O（log |D |•n），客户端为O（log |D |）。

我们的解决方案的核心是从在大（指数大小）域上以分布式方式计算DP直方图的问题减少到在小域上计算与输出直方图中非零的数量成比例的匿名直方图的问题。为了实现这一点，我们利用加密技术对遗忘伪随机函数（OPRFs）[54,64]进行分布式评估，这使两个计算双方能够将索引从直方图域转换为伪随机域，允许在隐藏实际值的同时进行聚合。

我们还开发了新的分布式DP协议，用于计算匿名直方图，其中服务器不能访问清晰数据中输入的索引。我们的第一种技术依赖于密文的复制和再随机化，而我们的第二种替代技术建立在一个安全的双服务器实现上，如Boneh等人的一个。[16].

除了对协议的渐近分析之外，我们提出了通信和计算成本的实验评估，并将我们的协议与使用混乱电路[77]的基线进行比较。我们的结果表明，我们的协议规模随着当事人数量的增加，由于他们的输入数量的线性复杂性。对于适合单个密文的10亿用户和域元素，我们可以只使用1.08 ms的服务器计算和每个用户服务器之间339字节的通信来计算DP直方图。同时，每个用户只需执行0.46 ms的计算，并在一条消息中通信192个字节。

相关工作。Bohler和克施鲍姆[14]提出了一个用DP计算近似重击球者的双方协议。与我们的基线（第3.1节）一样，他们的协议使用了通用的MPC。事实上，他们具有t = n的算法1在功能上与图14等价，t对精度与性能的权衡值较小。相比之下，我们的主协议通过允许两个服务器学习关于输入的额外私人信息，比混乱的电路基线优越了一个数量级。

我们在第4.2节中使用的克隆技术类似于在[24]中使用的假用户技术。一个关键的区别是，即使是真实的用户，他们也必须使用类似RAPPOR的程序随机化输入(cf。[34])，也就是说，翻转每个桶中的位（参见[24，算法2]）。对于匿名直方图来说，这并不容易做到，因为这里的“桶”是不能从每个（加密的）输入中确定的多样性。因此，他们的协议不能应用于我们的设置。

2背景和模型

2.1隐私

我们使用supp (U)来表示对一个分布U的支持。我们也写出pU (x)来表示U在x处的概率质量。对于k∈N，我们写了$U^{\bigstar k}$表示来自U的k个独立样本的和的分布，即U的k个卷积。为了方便起见，我们为一些∈R写一个+U来表示+X的分布，其中X∼U。我们有时也会用一个随机变量来代替它的分布，反之亦然。

冰球棍在分布U之间的差异，U′ 是

$$\text{德}(U/\!/U\text{‘}) := \sum_{x\in \mathrm{supp}(U)} [pU(x) - e^{e}\cdot pU'(x)]+,$$

其中，+:= max{g, 0} .

我们说，两个分布的U，U′ 为（c、6）-不可区分，记号为$U \equiv_{e,6} U'$，如果，如果，如果，如果（U//U ‘），如果，（U’ //U）≤6。我们考虑两个数据集X，X′ 如果是X，则为相邻的′ 通过改变单个用户在X中的贡献而得到的结果。

差异的隐私。如果对于每一对相邻数据集，一个函数f称为（c，6）差分私有（或（c，6）-DP）[31]′ 它认为，f(X) ≡e,6f(X′).

上述邻近的概念在文献中被称为替代DP。作为证明的一部分，我们将使用添加/删除DP的概念。这是由我们说的X来定义的′ 邻居X，如果一个是通过删除一个用户从另一个到达的。我们将使用添加/删除DP意味着替换DP这一事实。[1]然而，我们没有为整个协议提供添加/删除DP保证，因为我们的协议中的服务器的视图包括用户的数量。 我们使用以下概率分布族。泊松分布，记为Poi (7)，是具有质量函数exp（-7）7x/x! 的离散非负分布。负二项分布，记为NBin（r，p），是离散的非负dis-

惩罚与质量函数给出的$\binom{x+r-1}{x}(1-p)^r p^x$.DIS
离散拉普拉斯分布，记为DLap（入），是离散的分布分布
使用质量函数∝exp（-|x|/入）。我们将使用（离散的）

拉普拉斯机制，即，事实上，添加一个噪声样本来自
DLap（入），与入= Δ/e，对敏感度的结果-Δ（离散）
查询提供了（e、0）-DP。截断的离散拉普拉斯分布区
表示TDLap（入，t），是{-t，...，t}上的离散分布
使用质量函数∝exp（-|x|/入）。我们将使用添加的事实
从TDLap（入，t）的噪声样本，与入= Δ/e，结果
Δ查询提供的敏感性(e，$2e^{-(t-\Delta)e/\Delta}$) -DP.这是
从下面的尾部绑定，我们使用在整个pa-

per：对于X∼DLap（入），它认为Pr [|X |≥s入]≤$2e^{-s}$.因此
设置t=Δ+Δ/elog（2/6）提供了（e，6）-DP。我们用这个

1如果f是（e，6）-添加/删除DP，则f是(2e，（1+exp(e)）6）-替换DP。

在我们需要有界噪声样本的情况下的机制。截断的移位离散拉普拉斯分布，记为TSDLap（入，t），是在{0，...，2t}上的离散分布，质量函数为∝exp（-|x-t|/入）。在这种情况下也有类似的结果：将来自TSDLap（入，t=Δ+Δ/elog（2/6））的噪声样本添加到Δ查询提供的（e，6）-DP的敏感性结果中。我们在需要正噪声样本的情况下使用这种机制。

2.2安全

同态加密。同态加密（HE）是一种允许对加密数据进行计算的原语。在我们的构造中，我们只使用具有函数保密性的加性HE方案，用AHE表示。我们的主要构造依赖于其加性同态变体中的ElGamal加密（完整版本[9]中的图13）。

车库电路。混淆电路[77]是一种安全的双方计算的通用方法，它能够对任何可以由布尔电路表示的函数进行安全的评估。这是一个一轮协议，其中一方，即参与者，准备一个评估电路的编码，称为混乱电路（GC），并将其发送给另一方，即评估器，后者只能在具有相应混乱编码的一组输入上评估GC。收集器提供其自己输入的编码，双方运行一个协议，使评估器能够获得其输入的编码。

不明显的伪随机函数（OPRF）。伪随机函数（PRF）[44]是一个键控函数fk这样的输出fk(x)与随机是无法区分，即使输入x是已知的，只要密钥K是秘密的。一个被遗忘的PRF [54,64]是一种PRF，它有一种评估它的机制，使持有密钥K的一方不学习输入x，而提供输入x的一方学习fk(x) .

在我们的协议中，我们使用PRF FK (x) = H $(x)^K$由Jarecki等人介绍。[54]证明，当H被建模为随机预言时，这个函数是伪随机的。

2.3设置和威胁模型

我们论文的目标是在不信任任何一方的情况下，计算许多客户持有的输入的DP直方图，而不信任任何一方。我们通过在两个服务器上分配信任，并让它们使用交互式安全计算协议来计算直方图。这些服务器被认为是半诚实的，也就是说，它们遵循协议的步骤，此外，它们是非串通的，彼此之间不共享或接收任何信息。

我们要求我们的协议的输出来保证（c，6）-DP。然而，由于DP的最初定义假设了一个中央的、受信任的管理员，因此它不会立即推广到多个方面。贝梅尔等人。[8]将DP的概念扩展到多方设置，要求被对手破坏的每个政党子集的视图是DP；他们的工作集中在信息理论上，没有计算假设的设置。米罗诺夫等人。[65]引入了计算DP（CDP），它允许一个有计算边界的对手，并已在最近的工作[41,73]中使用。他们最强大的隐私概念，SIM- -CDP，要求所讨论的协议安全地（在MPC[45,59]的理想/真实模拟范式中）实现一个功能

提供DP。这意味着，MPC协议的分布式执行不会向任何一方透露任何信息，而计算的输出还提供了DP属性。正如他们所显示的，这是比仅仅更有力的保证
要求各方在执行过程中的观点是DP。

在MPC文献中，多项作品[50,62,70]探讨了DP泄漏的概念。这放宽了常规的MPC保证，即没有一方可以学习除输出之外的任何东西，通过允许参与者学习额外的信息，但强制要求所提供的额外信息是DP。形式上，这是通过捕获协议执行期间显示的附加信息作为泄漏术语来建模的，该术语提供给安全证明中使用的模拟器。这允许针对相同的功能比较不同的协议的泄漏，这可能会有很大的不同。特别是，它允许对泄露的信息进行更细粒度的控制，超出了DP。

我们的安全定义遵循相同的范例，并要求协议明确定义它们的泄漏L，并与输出一起显示。实现功能F的协议对于泄漏L是安全的，如果它计算F，并且可以从（F，L）模拟其中的视图。我们要求F和L被共同定义，以便定义它们的关节

在一个函数F^中的分布。

定义2（视图）。设Π是一个带有输入的两方协议

从X1×X2开始，然后视图(x^{Π}_1，x_2)表示乙方的观点
在使用输入x执行π期间1来自P1和x的∈X1∈ X2

从P2。该视图包括接收到的所有消息，以及所有收到的消息
在执行过程中抽样的随机数字（见Goldreich [45，第7.2节]）。

定义3（带泄漏的功能）。让F^=($F^_1$,$F^_2$) = ((F_1，L_1)，(F_2，L_2))是一个两方化的功能，从X1× X2 .让F = (F_1，F_2)和L = (L_1，L_2) .我们说的是一个两方协议Π
安全地实现∈泄漏L，如果每个b∈{1,2}有
提出了一种概率多项式时间算法西姆让所有的人
x_1∈ X1，x_2∈X2，其输出值为(Simb (xb,F b(x_1，x_2))，F(x_1，x_2))是
我们的F^具有泄漏的功能(x_2). Π(x_1. x_2)).

请注意，这个定义并不要求泄漏被解除
由Π精确计算，如果我们要求
一个F^的安全计算。这也意味着我们会怎样做
不排除学习L的可能性1和L_2一起可能
泄漏输出。因此，我们的要求是基于这个原因
不与对手勾结的政党没有透露任何事情
他们被泄露到另一方。这包括通过任何进一步的方法
采取的行动。然而，我们确实允许，就像在经典的MPC中一样，每一方与对方分享他们的输出，或在后续使用它
计算

恶意客户。虽然这项工作的重点是构建一个分布式聚合协议，保护客户端的隐私，另一个担心实际部署可能是恶意客户端提供不正确的输入倾斜输出和呈现它无用，或与两个服务器之一揭示诚实的客户的价值。在这种情况下所采用的主要方法是通过增加零知识，将客户的贡献限制在一定允许的范围内

证明来自客户端的[47]允许聚合器验证客户端的输入而不学习任何进一步的信息是有效的。防弹[20]等技术使客户端能够为其输入的范围生成一个证明，这可以由任何其他方进行验证。Prio工作[26]的方法，使范围证明，利用两个非合并的验证器，以实现更好的效率。本文介绍客户的范围证明，并与我们的结构集成，以防止恶意客户端，是未来工作的一个有趣的课题。

3个目标功能和基线

在本文中，我们的目标是实现一个分布式版本的Korolova等人的作用机制。[57]，也由Bun，Nissim和Stemmer [19]在不同的背景下引入，有时被称为基于稳定性的直方图。给定一个数据集I=(indi)i ∈ [n]对于来自一个大域D的索引，机制(i)构建I的直方图H，（ii）将DLap（2/c）噪声添加到H的每个非零条目中，（iii）删除值低于阈值T = 2 log（2/6）/c的条目，以及（iv）释放得到的直方图。选择这个阈值是为了使释放一个真计数为1的索引的概率以6为界。每个客户端可能贡献更大值的变体vali∈[1，。..，Δ]，因此输入是一个集合I = (indi，vali)i ∈ [n]可以通过添加DLap（2Δ/c）噪声和设置T = Δ + 2Δ log（2/6）/c轻松地处理。

3.1通用MPC解决方案

一个直接的解决方案是应用通用的双方计算（2PC）
在前面所述的中央DP机制的两个服务器之间
在…上面客户端秘密在两个服务器上共享他们的输入，然后服务器使用通用的2PC，例如，使用混乱电路来实现上面描述的目标功能。回想一下，混乱的电路协议要求我们将计算出的函数表示为一个布尔电路。因此，使用带有太多输入相关操作的朴素编码会破坏电路的大小，从而增加计算成本（电路中与门的数量是线性的）。完整版本[9]附录中的图14展示了我们的目标功能的数据无关算法，该算法依赖于众所周知的大小为O（n logn）[1]的排序/排列网络，得到一个大小为O（log |D |•n logn）的电路[2].该解决方案的灵感来自于基于专用集交集[53]的混乱电路的排序-比较-洗牌方法。在我们的实验评估中，我们使用它作为基准的结果协议的属性，在下面的定理中被捕获。如上所述，我们的解决方案的一个独特的方面是服务器成本为O（n•log（|D |））。

定理4。考虑n个客户端，每个客户端都持有一对内迪∈D，vali∈[Δ]。有一个一轮的安全协议，依赖于两个非合并服务器P1，P2为P1来获得输入数据的DP直方图，使用1∞-error O（Δ log（1/6）/c）。通信和计算为O（log |D |•n logn），客户端为O（log |D |）。

请注意，定理4中的通信和计算成本比定理1中的成本大一个log n因子。

[2]在实践中，排序网络的大小为O (n $\log^2$n)的使用是因为其更好的混凝土效率[7]。

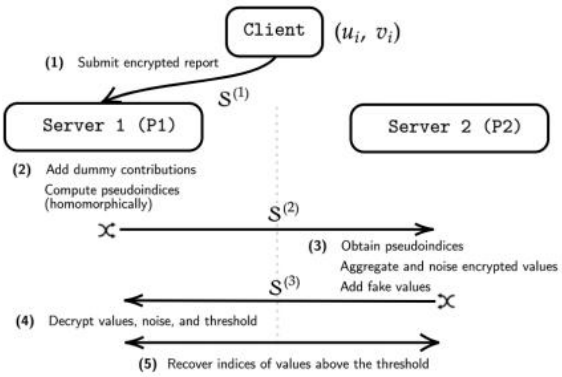

图1：我们的主要协议的高级流程。

.23洗牌DP

另一个可能的基线是使用来自洗牌DP文献中的协议。回想一下，DP [11,23,33]的洗牌模型是一个
介于DP的局部模型和中心模型之间的中间模型，
其中，客户端将消息发送到一个受信任的洗牌程序，后者在将所有用户的消息发送到分析器之前，会将这些消息随机地排列在一起。要求是分析器的视图（或相当于多组消息）需要是DP。通过实现安全的洗牌（例如，通过洋葱洗牌），可以在一个双服务器设置中实例化洗牌DP模型。

直方图查询在洗牌模型[5,23,24,36,39,40,42,43]中得到了很好的研究。不幸的是，虽然已知Oc（log（1/6））是可以实现的[5,42]，但已知的协议受到影响 从…通信的复杂性随着

Ω_c 日志（1/6） 其中|D |表示域大小。这在我们感兴趣的环境中非常大，其中|D |≫n；因此，我们不能在实验中使用它作为基线。

4技术概况

回想一下，对我们的问题的输入是一个集合，我∈[n]客户持有的（索引、价值）对。

如上所述，我们的协议将输入数据的DP视图泄漏到每个非合并服务器。直观地说，我们的协议显示，除了输出外，还有一个DP匿名的直方图
我∈[n]到其中一个服务器，和一个DP匿名直方图的{vali}
i∈[n]到另一个。回想一下，一个匿名的直方图对应于每一个i > 0出现i次的值的数量。个人输入的隐私(以及小的输入的隐私组)因此在DP的精确意义上被保护，而保护输入作为一个整体（在基于标准模拟的MPC的意义上）被牺牲为效率，如上所述。

4.1主协议说明

协议步骤的大纲。图1显示了对我们的主协议的高级描述，它涉及到服务器P1和P2之间的四个交互步骤，如下所述。

步骤(1)：客户端向P1提交一个加密的报告，构成一组S (1)的密文，在P2持有的密钥下加密-值被直接加密，而索引被散列，然后
加密由于（ElGamal）加密的特性，P1可以操作中加密的报告$S^{(1)}$为了同态，即，在没有事先解密的情况下，(i)将哈希索引H（indi）随机化为伪索引H（indi）^{K}，以及（ii）复制和重新随机化加密。

步骤(2)：使用这两个操作，以及模拟额外的虚拟贡献，P1构造了一个集合S (2)对包含原始客户端贡献集的（伪索引、值）对的加密。第二组件在P1具有密钥的AHE方案下进行加密，然后另外使用P2具有密钥的语义安全加密层，以保护这些值不受P1的影响。P1发送$S^{(2)}$以随机的顺序排列到p2。S (2)中的虚拟贡献值为0，这样它们就不会影响最终的直方图估计。

步骤(3)：P2解密中的密文$S^{(2)}$，并按它们的第一个组件对它们进行分组。请注意，这只揭示了每个索引的多重性，因为这些索引是伪随机的(它们被编码为H（indi）^{K})，并且这些值是加密的。然后P2同态地相加值，并将得到的值集返回到P1，并以随机顺序返回[Δ]（加上拉普拉斯噪声）中的虚拟加密；设S (3)是这个集合。虚拟值的目的是确保P1能够在清除中对P1同质聚集的值进行解密和阈值（阈值T），同时保留DP。

步骤(4)：伪指数倒置，P1学习直方图。

上面的描述只是一点简化，因为我们不能“倒置”伪指数。相反，每个客户端也发送一个加密- 在P1、P2的密钥下加密的索引(表示为双在图5中)；这些加密的索引与上述的伪索引和值一起传递，并且只对通过阈值的索引进行完全解密。

假贡献和DP。请注意，有两个步骤可以注入虚拟贡献：步骤(2)和步骤(3)。在这两种情况下，虚拟贡献的分布都是仔细选择的，以确保对方可以了解输入的量，观察各自步骤中的流量，有DP的意义。这导致了计算/通信成本和隐私之间的权衡。

具体地说，在步骤(3) P2学习输入指数集合{indi} i∈[N]的匿名直方图（即直方图的直方图），定义为其条目的直方图H你好包含输入中具有多重性i的索引数。但是，这是有泄漏的$S^{(2)}$揭示了输入中每个索引的多样性。不幸的是，这使得我们的协议不是DP(例如，如果对手知道除了一个指数之外的所有指数，那么它可以从H推断出剩余的指数是否与其已知的指数一致。)

如上所述，我们通过让P1插入虚拟贡献来克服这个问题$S^{(2)}$，除了与输入所对应的那些输入之外。正如我们将在下面解释的，仔细选择虚拟贡献的分布可以确保$S^{(2)}$现在只泄露了一个DP匿名的直方图。请注意，步骤(3)中的情况是类似的，如在这种情况下，P2插入了虚拟贡献，以确保P1的协议视图的DP。这种方法的一个核心挑战是在平衡隐私和通信之间的权衡方面：虚拟贡献有助于提供有意义的DP

虽然有保护作用，但可能会破坏通信。我们解决方案的一个主要组成部分是有效地做到这一点的机制，我们将概述。我们通过从一个更简单、效率更低的方法开始，然后发展到更复杂、更高效的方法来构建我们的解决方案。

.24个通过复制得到的匿名直方图

在本节中，我们提出了两种不同的协议，以在输入分布的互补假设下实现DP。我们的混合协议将对应于按顺序运行这两个协议。（在完整版本[9]的附录E中，我们提出并评估了一个完全不同的协议，它基于私人的重量级球员；只有在少量重击球员的设置下，这在渐近和数字上都具有通信优势。） 更具体地说，对于一个阈值T，第一个协议（过多重性噪声）只向用户的贡献提供DP多重性最多是T，而第二种协议，基于重复的噪声，保护了至少有T的多重性的输入。

每多重性概念：一种针对小多重性的有效协议。生成DP直方图的一个标准方法是在每个条目中添加适当的缩放的（离散的）拉普拉斯噪声。要在我们的设置中实现这个想法，P1将必须做到添加Oc,6(D)假贡献（Oc,6(1)每个可能的索引）。由于P2观察到一个匿名的直方图，所以只要对一个多重性H的直方图就足以产生噪声了，这是一个轻微的优化你好计算具有多重性i的伪指数的数量$S(2)$.在我们的设置中，P1可以通过添加虚拟索引的贡献来实现这个机制（来自与原始域不相交的域I）来实现这个机制；用单个虚拟索引添加i贡献相当于向i匿名直方图条目添加值为1的噪声你好.由于H可以有多达n个非零项，并且每个项都必须添加噪声，因此P1需要addzi∈[n]Oc,6(i)=Oc,6(n^2)有不同的这类贡献，以确保这一点$S(2)$是DP。然而，如果我们可以假设没有一个伪指数的多样性超过一个阈值T，i。e.，那个∀i ＞ T：你好=0，然后提高到多重性T就足够了，开销是Oc,6(T2)；这显然对大的T是不受欢迎的。

复制：一种有效协议。

请注意，P1并不局限于模拟虚拟贡献： 由于ElGamal加密允许重新随机化，所以P1可以忽略- 疯狂地生成一个对（indi，0）的加密 （indi，瓦利），既不学习indi，也不学习vali。我们将利用这种“复制”能力来构建一个协议。

下面的观察结果是至关重要的。考虑一个输入数据集D和另一个数据集D'与D相同，除了没有客户机1的数据，以及相应的匿名直方图H、H'对于各自的集合I=我∈[n]和$I' = \{indi\}$我∈$[2_{..n}]$的指数。

请注意，这些数据集在添加/删除的意义上是相邻的（第2.1节）。现在，假设x是ind的多重性$_1$在I中，和注意到H，H'不同的是，只在两个相邻的项x，x-1，作为删除ind_1从I减少具有多重x的指数的数量1，同时增加指数的数量x-1由一个。更准确地说，它认为$Hx = H+1$，$'_x$ $Hx -1= H'_{x-1}-1$，和∀g∈{x，x-1}，$Hg= H'_g$.(请注意，这不如a更一般直方图，其中相对于相邻数据集的变化可以发生在任意的桶中，尽管ℓ_1在这两种情况下，灵敏度都以2为界。)

考虑当我们从I中复制每个索引，从一个分布U中随机抽样的次数时，H是如何变化的。下面的算法Dup (H)将描述，该算法返回修改后的直方图有给定H：

Dup (H):

有=∅◃ 为空的直方图

对于g∈Dom (H)

重复H_g乘以

样本a～ U★g

$H^d_{g+a} \leftarrow H^d_{g+a} + 1$

返回有

注意，该算法迭代原始直方图H的每个项g，通过～“移动”每个对入口g的贡献U★g的复数形式因此，a对应于一个具有多重性g的索引的附加副本的总数，其中它的每个g实例都是重复的X～U次。

为了满足DP，应该选择D来满足Dup (H) ≡c,6Dup (H′) .由于H和H在条目x-1和x上相差1，并且在其他地方相等，这可以归结为$ax -1 \equiv c,6 1 + ax$，其中$ax -1$～ U★(x -1)，ax～ U★x.这几乎与添加的机制的条件相同U★x噪音是DP(它只是取代ax -1～ U★(x -1)和ax -1～ U★x).事实上，我们证明了一些众所周知的分布U，如负二项分布——已经用于DP——满足我们更严格的条件，假设x ＞ T。注意，后一种假设是必要的：如果x = 1，那么该条件显然失败为ax -1=总是0，而1 + ax≥1。为了实现DP，我们必须有ax -1>为0，至少带有Oc,6(1)概率因此，每个项目所需的预期重复数为Oc,6(1/T)，产生为Oc,6(n/T)完全一样的东西

一个混合协议：两全其美。我们依次组合重复和每多重噪声协议获得混合协议。为此，我们首先将每多重性噪声添加到一个预定义的阈值T上，然后将复制协议应用于结果集。这就留给了我们选择T的任务。由于我们的目标是最小化由P1插入的虚拟贡献的总数，因此这可以归结为优化Oc,6(T2+n/T)（在实践中我们是通过数值来执行的），对应于这个混合协议的开销。

正如我们在后面的实验中所演示的，分别使用TSDLap（•）和NBin（•）分布实例化的混合协议产生了一个实际的协议，但我们可以做得更好。接下来，我们介绍一个导致最终协议的优化。

一个改进的协议。注意，一旦阈值T固定，混合协议将DP证明减少为两种情况。假设，如上为受保护用户索引的次数。如果是x≤T，则添加拉普拉斯噪声提供DP，如果是x ＞ T，则重复提供DP。现在让我们考虑混合协议为多重性x = T-1增加的噪声量，即，当x很大，但不够大到足以被重复保护时。在混合协议中，具有多重x的输入只受到每多重噪声的保护，即使它们

通过复制来实现DP。这是不令人满意的，因为在这种情况下，重复会导致“浪费”的通信开销，而没有改善隐私。为了解决这个问题，我们将引入一个仔细校准的泊松噪声来补充重复。

在一个高水平上，我们引入了一个中间状态（T，T‘）。对于T的多重性，我们将使用每桶噪声；对于大于T′，我们将使用（适当校准的）重复。接下来，我们将描述如何在（T，T‘）中以多重性来保护输入。 设x是（T，T‘）中的多重性。重复后，新的多重性x+U★x在间隔中“展开”[x，∞）。特别地，这意味着在每个多重性j∈[x，T‘]中添加Oe，6(j)噪声（就像每个多重性噪声所做的那样）是一种过度利用。相反，额外的Oe,6(j)噪音可以像x一样被传播出去+ U★x是分散。我们通过在每个多重性j中增加一个Poi（7j）量的噪声来实现这一点，其中7j的精心挑选因素渐近地，这似乎改善了所需噪声对6的依赖，并进行了实验所示的实际改进。

对这种方法的分析很大程度上是受到Feldman等人对洗牌DP的工作的启发。[36].特别的是，我们认为补充泊松噪声创建（随机数量）“克隆”的x或x-1。利用泊松分布的性质，得到了
这些克隆体的数量也遵循泊松分布；我们
表明DP只要达到预期的克隆数量
足够大。然后用这个条件来选择我们的选择7j在实验的理论和数值上。

与之前的DP匿名直方图工作的关系。

在我们继续之前，让我们提到一下，有几个作品已经考虑到-
解决了在中心模型[2,12,51,52,61,72]中释放DP匿名直方图的问题。事实上，最优的中央DP算法[61,72]也为x ＞ T和x≤T添加了单独的噪声，类似于我们的混合协议。然而，我们强调，我们的设置更具挑战性，因为我们不能直接向H添加噪声；我们可能只创建具有新索引的假人或复制现有的索引，而不知道输入的（加密的）真实索引。这就是为什么我们需要使用新的噪声方案来实现我们的目的。另一项相关的工作是Ghazi等人。[41].虽然这项工作不是用于计算DP直方图，但它也存在匿名直方图的泄漏，并使用每多重噪声使匿名直方图DP。但是，如第4节中所解释的那样。2，这种技术单独对我们的设置是不够的，我们必须开发一些额外的技术（复制和克隆）来使协议实用。

5我们的协议

在本节中，我们将按照第4节中概述的高级方法，详细描述了DP稀疏直方图的主要协议。我们将目标功能（图4）分为两部分：我们在5.1节开始，描述一个阈值功能（图2）和协议（图3），该协议允许两个服务器揭示一组通过特定阈值T的加密值中的DP值。在第5.2节中，我们在更大的协议（图5）中使用该功能来计算私有直方图。

公共参数：
噪声参数入、t和阈值t。
具有公钥PK1的AHE方案。
输入：
P1： SK1，与PK1对应的密钥。
P2: Ciphtertexts(wi)i ∈[n]，其中每个wi的形式为Enc（PK1，vali）。
功能：
(1)对于i = 1，。.., n:
(a) $5_i^{(1)}$, $5_i^{(2)}$ ←RTDLap(入, t), 5i ← $5_i^{(1)}$+ $5_i^{(2)}$
(b) vali ← $\begin{cases} \mathrm{val}_i + \xi_i & \text{if va 李 + 5i } \geqslant \text{T,} \\ 0 & \text{oth e rwi se} \end{cases}$
(2)返回（vali）i∈[n]到P1。

图2：目标功能的F阈值。

在下面的两个小节中，我们使用相同的结构：首先，我们描述目标功能，然后是我们的协议。然后，我们定义了我们的协议的泄漏功能（参见定义3）。在我们的论文[9]的完整版本中，我们随后证明了我们的协议在给定的泄漏下安全地实现了目标功能，并表明组合功能（目标功能+泄漏）的输出提供了DP。

在完整版本的附录E中，我们还描述了另一种基于私有重击者协议的私有直方图的方法，并将其与本节中我们的主协议进行了比较。

5.1阈值协议

在本节中，作为一个热身，我们将描述在图1中的步骤(4)和(5)基础上的阈值协议。它给出了图2的理想功能。p2的输入是同态的
加密的密文（wi）i∈[n]，而P1持有相应的密钥。功能首先，在步骤（1a）中，样本噪声从一个截断的中心离散拉普拉斯分布，并被添加到每个解密值。然后，在步骤（1b）中，它将所有低于阈值T的值设置为零。最后，在步骤(2)中，将阈值返回给双方。

我们用来实现阈值化功能的协议是
如图3所示。在该协议中，有两个泄漏源。
首先，每个方都保留自己的噪声值5i，添加到每个条目i∈[n]。这意味着双方可以在局部计算出比理想的功能输出噪声更小的输出版本。因此，我们必须在泄漏中包括各方各自的噪声份额。第二个泄漏来源是P1在阈值前只添加P2的噪声来学习所有值。

在下面的正式描述中，我们省略了双方的输入
为了可读性，e。g.，我们写$L^{P1}_{threshold}$表示$L^{P1}_{threshold}$ (x^0, x^1).

定义5（泄漏Π threshold ）.让$5_i^{(1)}$, $5_i^{(2)}$发出声音
分别在图2的步骤（2ii）和（1a）中生成的样本。然后我们定义了Π threshold的泄漏:

$$\mathcal{L}^{P1}_{threshold} = \left\{ \left(\xi_i^{(1)}\right)_{i\in[n]}, \left(\mathrm{val}_i + \xi_i^{(2)}\right)_{i\in[n]} \right\},$$
$$\mathcal{L}^{P2}_{threshold} = \left\{ \perp \right\}.$$

公共参数：
噪声参数入、t。
阈值T > 2t。
具有公钥PK1的AHE方案。
输入：
P1： SK1，与PK1对应的密钥。
P2：密文(wi)i ∈ [n]，其中每个wi的形式为Enc（PK1，vali）。
协议：
(1) P2:
(a)对于每个i∈[n]，使用同态加密属性向加密值添加噪声：

$$S(1) \leftarrow (\mathrm{Enc}(PK1, vali + 5i))i\in[n]$$

其中，5i←TDLap（入，t）。
(b)发送S (1)到P1。
(2) P1:
(a)为每个记录$w'_i \in S(1)$收到P2：
(i)解密时间$'_i$ ←Dec（SK1，w）。$'_i$
（ii）样品$5'_i$ ←TDLap（入，t），和计算$val'_i{}' \leftarrow val+5'_i{}'_i$.
（iii）如果$'_i{}' < T$, $val'_i{}' \leftarrow 0$.
(b)集S (2)←（val ‘）i∈[n]$'_i$.
(3) P1输出S (2)。

图3：使用泄漏L阈值实现F阈值的协议ΠthresholD.

带有泄漏的功能被定义为联合分布
（阈值，$L^{皮}_{th}$回复），表示为F^threshold.

定理6。协议Πthreshold在图3中，设置了入1=入，从而安全地实现了阈值从图2与泄漏L阈值=$(L^{P1}_{threshold}, L^{P2}_{th}reshold)$在定义5中。

证明内容见完整版本[9]的附录B.1。

定理7。让入= 2Δ/e和t = Δ +入log（2/6）。然后，对于i∈{1, 2},F^threshold=（阈值，$L^{皮}_{th}reshold$）是一个（e，6）-DP函数在数据库上(valj∈ [Δ])j ∈ [n].

证明。语句来自截断的拉普拉斯机械
一旦我们必须知道$5^{(1)}_{i}$的(报告。⊥),P1
(报告。P2)观察一个参数为入的拉普拉斯噪声直方图。
注意，T的阈值是后处理。□

5.2私人稀疏直方图

我们现在描述了私有稀疏直方图的主要协议。我们在图4中给出了对目标功能的正式描述。
它紧跟着第3节中的高级描述，与
主要的区别是，我们明确地将噪声项5i分解为两个组件，其中一个组件通过上一节中的阈值化协议泄露给每个助手服务器。

我们的主要协议如图5所示。它遵循图1中的大纲。
在步骤(1)中，每个客户端i首先从它的（indi，vali）对中准备三个密文：一个包含indi哈希的加密，它将用于获得indi的OPRF值。第二个是加密indi（没有散列）。这是用于恢复的

公共参数：
DP参数e，6，灵敏度Δ。
噪声参数入= 2Δ/e和t = Δ +入日志（2/6）。
阈值T = Δ + 2t + 1。
输入：
客户机：索引值对，vali)i ∈ [n]
功能：
(1)出租$(ind'_j)j \in [n']$表示输入中的唯一索引。对于每个j∈[n‘]：
(a)示例$5^{(1)}_j$, $5^{(2)}_j \leftarrow_R$TDLap(入, t).
(b)计算$5j \leftarrow 5^{(1)}_j + 5^{(2)}_j$和

$$val'_i = \begin{cases} \perp & \text{if 司} + \xi_i < \tau \\ s_i + \xi_i & \text{oth erwise,} \end{cases}$$

在哪里司=$^{z}_{\{j \mid indj = ind'_i\}}valj$.
(2)输出$'_j$, val) | j ∈ [n′], $val'_j{}'_j$ =⊥}到P1。

图4：目标私有直方图功能Fhist。

聚合后通过阈值的明文桶ID。最后，客户端使用P1的AHE公钥对其值进行同态加密，然后再次使用标准加密对P2的公钥下生成的密文进行加密。这允许P2同态地添加属于同一桶的客户端贡献，同时隐藏来自P1的值，直到它们通过外部加密层聚合。

在步骤(2)中，P1用其秘密的PRF键K对每个第一个组件进行指数化，以对P2隐藏明文索引。然后继续添加第4节中讨论的虚拟值，使用零的加密作为第三个组件，以确保虚拟值不会添加到聚合值中。我们在图6-9中描述了虚拟采样算法的细节。生成的密文集被打乱，然后发送到P2。

P2现在可以在步骤(3)中解密这三个组件中的两个。在步骤（3a）中进行解密后，

- h′i 是equalto $H(u'_i{}^{K})$va借给恩克斯(PKAHE′i , d),
- w′i similarly equ,i

′ ′
i 对于一些索引值对（u，d），要么由客户端贡献，要么由P1作为一个假人添加。

解密后，P2同态地聚合共享相同第一个组件的三元组的所有第三个组件，任意选择第二个组件中的一个。现在请注意，在这一点上，聚合桶的数量与P1的视图没有区别地是私有的，因为P1确切地知道在步骤（2a）中添加了多少虚拟桶。为了解释这一点，P2必须添加额外的虚拟桶，这是在步骤（3d）中对采样桶的调用中完成的。之后，双方对聚合值调用Πthreshold，获得高于阈值T的明文值。我们已经选择设置T为大于Δ + 2t1为了保证在步骤（2a）中由P2添加的假人即使在添加了两个TDLap样本后也总是低于T。请注意，如果我们允许假人有概率超过阈值，这可能会得到优化2^{-G}对于一个统计安全参数G。我们把这个优化留给未来的工作。

公共参数：

具有发电机g的素数阶q的组G。

直方图索引域U = {0，...，2^d-1}，值域V = {0，。..， Δ} .虚拟索引域I与U∩I=∅。

随机神谕H： U∪I→G。

ElGamal公共加密密钥PK = PK1 • PK2，pk‘∈ G.公用加密密钥普卡赫对于可加性同态加密。用于语义安全加密的公共加密密钥pk‘’.

DP参数c=cleakage+ c计数， 6 =6泄漏+6计数，敏感性Δ。

自由噪声参数T、T′ 通过第6.2节中的网格搜索进行选择。

已确定的噪声参数入1=2Δ/c计数， t1≥Δ+入1日志（2/6次），入2=1/cleakage， t2≥入2日志（1/6泄漏），阈值T = Δ + $2t_1$+ 1.

输入：

客户端i：一个索引值对（ui，Ui）∈U×V。

P1： ElGamal密钥SK1∈Z_q是PK1的密钥，附加的HE密钥SKAHE对应于普卡赫，一个秘密的PRF密钥K←Z_q

P2： ElGamal密钥SK2，斯克∈Z_q，其中SK2是PK2的密钥，以及斯克是秘密的pk‘、解密密钥斯克“顺流而上pk‘’用于语义上安全的加密。

协议：

(1)每个客户端都计算hi←H（ui）和wi←EncHE（PKAHE，Ui）并进行加密。

（ai，bi，ci）←（EncElGamal(PK′，hi),EncElGamal(PK，ui),Enc（PK′′，wi））.

(2) P1接收来自所有客户端的密文$S^{(1)}$={（ai，bi，ci）} i∈[n].

(a)$S^{(1)}$←S (1)∪样本假人(T，T‘，裂解，6个泄漏，$S^{(1)}$).(b)

随机选择K-←RZ_q和设置$S^{(2)}$← ∅ .

(c)对于每个元组（ai，bi，ci）∈S (1)：

(i)打开包装(ct_1，ct_2)←艾。

（ii）a'_i ← (ct_1^K，ct) .$_2^K$

(iii)$S^{(2)}$← S(2)∪{（a、bi、ci）}。$_i'$

(d)洗牌S (2)，并将结果发送到P2。

(3)收到P2$S^{(2)}$= {(a，b，$c'_{i\,i\,i}$)} i ∈ [n′]从P1和计算：

(a)解密每个密文的所有三个组件如下

h'_i ←德克尔加马尔（斯克，a），$w''_{i\,i}$ ←12月(SK”，c)$'_i$

和设置$S^{(3)}$={ (h，b，w)}i ∈ [n′]

(b)通过定义，基于第一个组件来划分S (3)中的元组你好= {j | h'_j = h } .$_i'$

(c)对于每个唯一的值h'_i 在元组的第一个组件中$S^{(3)}$，同态地将所有第三个分量相加并进行选择

第二个成分之一是随机的。这将产生一个集合（按h排序）：$_i'$ $S^{(4)}$=((h$'_{i\,i\,i}$′′ i∈[n”]包含n′′ ≤ n′

形式的元组

(H(u)$''^K_i$,EncElGamal(PK，u),$_i''$ 恩克希(普卡赫, $\bigsqcup_{j\,\in\,Hi}$ U''_j))

(d)$S^{(4)}$← S(4)∪ SampleBuckets (PKAHE，入2，

t_2) .(e)洗牌S (4).

(4)出租（d'_i) i ∈ [n′′]和（w''_i) i ∈ [n′′]的第二个和第三个分量的有序集（以相同的顺序）$S^{(4)}$.P1和P2调用

Πthreshold从图3中，P1有输入SKAHE，P2有输入(w''_i) i ∈ [n′′]，设置T、入1和t1如上让（vali）i∈[n‘’]是

P1接收到的输出。此外，P2发送（随机阈值(d)）i∈[n′′]$'_i$ 他们P1。

(5)让V={{d'_i |i∈[n”]，|=0}为P1在上一步中获得的索引值对（具有加密的索引），不包括值为0的对。P1向P2发送以下集合，随机打乱。

D = {RandomizeElGamal (d) | (d，U) ∈ V} .

(6) P2计算{部分数据（SK2，d）| d∈D}，并以相同的顺序将其发送到P1。

(7) P1从步骤5和输出中恢复对D的洗牌{(DecElGamal(SK1，d)，U) | (d，U) ∈ V} .

图5：我们用于计算私有直方图的完整协议直方图。

参数：阈值T和T′′、隐私参数eleakage和6个泄漏，虚拟索引域I，一组消息S ={（ai，bi，ci）} i∈[n].
算法
(1)找到入3，t3，r，p,T′′和{7j}T ≤ j ≤T′′满足定理15与T，T′，e =eleakage/2，6 =6leakage/2(1 + exp (e))和6^=6leakage/2.
(2) S ← S ∪ SampleFrequencyDummies(入3，t3,T，I) .
(3) S ← S ∪ SampleDuplicateDummies (S，r，p) .
(4) S ← S ∪ SampleBlanketDummies({7j}j,T,T′′) .
(5)返回S。

图6：算法样本假人。

参数：阈值T，噪声参数入3、t3，虚拟索引域I
算法
(1) R ← ∅
(2)对于每个i = 1，。..,T:
(a)从TSDLap中随机抽取Ni(入3，t3) .
(b)为j = 1，。.., Ni :
(i)随机选择x′ ←RI .
（ii）执行图5i中的步骤1次，模拟具有输入（x‘，0）的客户端。将生成的密文添加到R中。
(3)返回R。

图7：算法采样频率假人。

参数： AHE公钥PK，噪声参数r和p，
虚拟索引域I，一组消息
S = {（ai，bi，ci）} i∈[n]，其中所有的bi都是可重新随机的ElGamal密码。
算法
(1) R ← ∅
(2)对于一个∈[n]：
(a)从NBin中随机抽取Ni。
(b)为j = 1，。..，Ni添加(ai、b'_j ,恩卡HE（PK、0））到R，其中为b'_j = RandomizeElGamal (bj) .
(3)返回R。

图8：算法样本重复复制假人。

参数：公钥PK、噪声强度7j，虚拟索引域I，阈值T、T′′.
算法
(1) R ← ∅
(2)对于每个T≤j≤T′′
：(a)重复Poi（7j）次
。
(i)随机选择x′ ←RI .
（ii）执行图中的第1步。5j次，模拟一个客户端的输入（x‘，0)。将生成的密文添加到R中。
(3)返回R。

图9：算法样本空白假人。

公共参数：
沙痂边界T。
噪声分布参数入1、入2，入3，t1，t2，t3、r、p和7i为我≥T
输入：
客户：指数值对I = (indi，vali)i ∈[n].
功能：
(1)设h0为第一个分量的匿名直方图的输入。也就是说，i^0表示不同的数量输入中出现恰好i次的桶。
(2)初始化N1，N3,N4∅。
(3)初始化H1←H0。对于每一个i = 1，。..,T:
(a)画Ni←RTSDLap(入3，t3) .
(b)N1 ← N1∪ {(i，Ni)}
(c) H_i^1 ← H_i^1 + Ni
(4)将H2初始化为一个空的直方图。每一个我。t.H_i^1=0，重复H_i^1时间：(a)画了一个←RNBin (i•r、p）。
(b) H^z_{i+a} ← H^z_{i+a} + 1
(5)初始化H3←H2。对于每一个≤≤′′：
(a)绘制N_i' ←RPoi (7i)
(b)N3← N3 ∪ {(i，N_i')}
(c) H_i^3 ← H_i^3 + N_i'
(6)让我′=($(ind'_i$ 我的∈[n′]]$'_i$ 是我首先分组的输入组件，将第二个组件相加，被打乱。
对于每个j∈[Δ]：
(a)从TSDLap中抽取Mj(入2，t2)
(b)N4 ← N4∪ {(j，Mj)}
(c)I′ ← I′ || ((⊥，j))大麻
(7)设置5(1)、5(2)←RTDLap(入1，t1)| I′ |，并让5=5(1) + 5(2).
(8)让我′′=((indj,valj + 5j)| j ∈ [|I′ |]，I'_j =(indj,valj)).
(9)定义Fh^{P1}_{ist}← {(ind，val) | (ind,val) ∈ I′′，val ≥ T}.
（10）定义Fh^{P2}_{ist}← ∅ .
（11）定义V⊥←{val |（ind，val）∈I“”，val < T}。
(12) L^{P1}_{ist}←(N1,N3，5(1)，V⊥)，L^{P2}_{ist}←(H3,N4，5(2)).
（13）将泄漏F^hist的功能定义为接头是distribution ($\mathcal{F}^{P1}_{hist}$,F^h^{P2}_{is}t)与F^h^{Pi}_{ist}=(**Fh**$^{Pi}_{ist}$，L^{Pi}_{hist}

图10：功能F^hist=（Fhist，Lhist）。我们展示了Π hist安全地实现泄漏。

最后，在步骤(6)-(7)中，P2和P1共同解密第二个步骤组件对应于阈值以上的值，其中允许P1获得这些值的明文索引。请注意我们需要在阻止P2链接解密的对其进行聚合的伪随机桶的索引。

接下来，在图10中，我们使用泄漏定义功能F^hist为我们的协议。而Fhist已经在图4中描述过了，我们将其包含在图10中，以使Fhist和L hist显式。F^hist的定义如下Π hist（图5）。这个直方图显示给双方在整个协议中，其中泄漏的组件对应于噪声注入样本直方图F^hist中的步骤(3)-(5)对应于调用

参数： AHE公钥PK，噪声参数入2、t。算法
(1) R ← ∅
(2)对于j∈[Δ]：
(a)样品Mj←TSDLap（入2，t）。
(b) For k∈[Mj]，生成虚拟记录(⊥，⊥，Enc (PK, j))，并将这些假人与R连接起来。
(3)返回R。

图11：算法采样桶。

Π hist的步骤（2a）样本，而F^hist的步骤(6)
对应于在Π hist的步骤（3d）中对采样桶的调用。笔记
那个H4不是直接透露给P1，而是在添加噪音后
每个条目，作为L的一部分$^{P1}_{threshold}$.步骤(7)-(9)对应于该呼叫到Π hist的Πthresholdin步骤(4)。请注意，所有的虚拟桶都通过步骤(6)中的P2，将低于T = Δ + 2t1+是1，所以没有假人桶将出现在Fh中$^{P1}_{ist}$.在步骤（12）中，我们定义了泄漏L hist双方。注意，我们将Πthreshold输出到P2

这里有部分泄漏，因为P2在Fh$^{P1}_{ist}$中没有任何输出。[3]

下面的定理建立了我们的协议的安全属性，即它实现了预期的泄漏的功能。然后，我们证明了输出和泄漏的联合分布是DP的。

定理8。协议Π（图5）安全地实现了Π（图4、10)和泄漏Π（图10)。请注意，图4中的c和6对应于ccounts和6计数在图4和图10中。

客户端与服务器合并。请注意，上面的安全性
当P1或P2与任何数字串谋时，参数仍然成立
对于半诚实的客户，因为客户不收到任何信息，和
对于任何输入，输出的分布都将是相同的
理想的和真实的模型。然而，如第2.3节所述，我们的工作
协议不能保护防止恶意客户端。特别是
请注意，我们的协议要求的值为H（indi）^{K}作为指数
从随机的处罚。现在考虑一个对手来控制
P2和客户端j。如果允许客户端偏离协议，
而不是提交一个加密的H（indj）作为它的第一个公司-
不，它可以为一个目标ind提交H（ind）$'_j$ $\frac{1}{2}'_j$.P2获得后
(a) i ∈ [n′]$'_i$ = (H（印地语）^{K})i∈[n‘]，它现在可以，对于所有的i∈[n’]，平方
a'_i 并检查a'^{2}_i= a'_i ′ 对于任何i ′ ≠ i .如果是这样的话，那么
对手知道$'_i$ ′ = H (indj)^{K}至少还有另一个
客户端提交指标。
通过让客户端证明它确实正确地评估了H，使用一个ZKP友好的哈希函数，可以避免上述攻击。[49].然而，这仍然允许恶意客户机通过提供毒结果
任意大valj。我们将协议的完整扩展留给恶意客户端，以备未来的工作使用。
5.2.1隐私保障。我们现在声明，对双方的输出结合的任何一个泄漏是DP。见附录B。3和B.4的完整版本[9]的证明。

[3]我们可以选择向双方显示输出的直方图菲斯特.从那时起
将需要额外的交流，我们只让P1学习
输出，并将在阈值化协议中学习到的噪声值p2视为泄漏。

引人注目9。(Fh$^{P1}_{ist}$, $L^{P1}_{h是t}$,Fht)$^{P2}_{is}$是（c，6）-DP。

最后10。(Fh$^{P2}_{ist}$, $L^{P2}_{h是t}$,Fht)$^{P1}_{is}$是（c，6）-DP。

5.2.2隐私分析H3.为了分析H3的隐私性，我们主要采用DP的添加/删除概念；最后，我们将使用该关系将我们的结果转移到替代DP上
在第2.1节中描述的两者之间。

让D是一个数据库和D′ 保持相同，但删除了第一个条目。设h3由D和计算出来H3′ 从…D′ 通过图10中的过程。回想一下，H3对应于算法6，由P1运行，对应于Π hist步骤（2a）中对样本的调用（图5）。让m1是内在的多样性1在D和m中$'_1$ = m1-1是在中的多样性D′ .

混合协议：每个多样性和重复性。我们将首先分析概述中描述的没有泊松补充噪声的混合方法。e.，案例T = T′ 我们在第4节中描述的最后改进的方法。如第4节所述，我们的分析包括两种情况，基于计数是低于T还是高于T′ .在前者中，隐私保证遵循被截断的拉普拉斯机制（如[30]），即：

引理11（低计数时的隐私性）。H3和H3′ 是（c，6），如果是m1≤T，入≥2/c和t≥1+入log（2/6）。

对于后一种情况，如第4节所述，我们可以证明这一点
1+ U★(T′ +1) ≡c,6U★T′ 其中U = NBin（r，p）。利用负二项噪声的尾界，我们可以选择具体的参数如下。

引理12（高计数者的隐私权）。设c，6∈（0,1）。H3和H3′ 是（c，6），如果是m1> T′ 其中T′ ≥3（1 + log（2/6））• ( ) $\frac{4}{\varepsilon}$（1 +日志（1/c））+$\frac{100}{\varepsilon^2}$，r =$\frac{3(1+\log(2/6))}{\varepsilon}$和p = $e^{-0.2c}$.

这些引理足以证明混合物的隐私性
无泊松补充噪声的方法。e.，如果我们取T = T′ 和7j=0表示所有j)。在这种情况下，我们得到以下定理。

定理13。设c，6∈（0,1）。如果T = T′ ≥3（1 + log（2/6））• ( ) $\frac{4}{\varepsilon}$（1 +日志（1/c））+$\frac{100}{\varepsilon^2}$，入≥2/c，t≥1+入日志（2/6），r=$\frac{3(1+\log(2/6))}{\varepsilon}$
andp = $e^{-0.2c}$，则图6中给出的算法是（c，6）-添加/删除DP。然后就立即发现它是（2c，（1 + exp（c））6）-DP。

此外，对于log（1/6）/c=o(=$^{1/3}$)，如果我们取T (=T′)= Θ($n^{1/3}$)，则在步骤(2)中生成和发送的虚拟消息的预期数为Θ($n^{2/3}$日志（1/6）/c)

以覆盖案例T < m1< T′ ，我们必须充分利用这样的机会7j.让尼加拉瓜～NBin（ri，p）和让Ti,j= P（Ni = j-i）（这是一个具有多重性i的消息被复制为具有多重性j的概率）。为了考虑将i增加1所做的变化，我们取最小的q，允许以下分解，用ai，j，pi，j和义、j分配

Ti,j=qiai,j + (1 − qi)yi,j, (1)

和

Ti+1,j=qipi,j + (1 − qi)yi,j. (2)

此外，定义pi是最小的p，如果A，B，C～Poi (p)是独立的，那么

$$P\left(\frac{q_iA+(1-q_i)C+1}{q_iB+(1-q_i)C}\right) > \text{expc} \leqslant 6. \quad (3)$$

该方法的主要隐私保证如下所述。如前所述，这个证明使用了来自[36]的技术，将每个补充泊松噪声视为一个“克隆”。完整的证明请见附录B。8个完整的版本的[9]。

最后14。如果所有j

$$7j \geqslant pm'_{1} \text{（上午，j+pm，j+ym，j）}_{111}, \quad (4)$$

那时H3和$H3'$（e，6）是无法区分的

最后的隐私保证在下面的定理中陈述，它是引理11、12和14的直接组合。我们的实验表明，该协议在实践中实现了对T = T的改进′通信中的情况。我们还提供了一个启发式的论证，即渐近性被改善了至少一个日志因子（1/6）$^{1/3}$在完整版本[9]的附录C中。

定理15。设e，6∈（0,1）。对于给定的T，T ‘，让入≥2/e，t≥1个+入日志（2/6），r=$\frac{3(1+\log 2/6))}{g(}$，p = $e^{-0.2c}$.此外，让

$$7j = \max(ai, j+pi, j+yi, j), \quad (5)$$

对于所有的j，然后选择T'′所以zj>T ‘’ 7j≤$6\hat{}$.

然后只要T ‘≥3（1 + log（2/6））• $\frac{4}{r}$（1 +日志（1/e））+$\frac{100}{r^2}$，那时H3是（e、6）-添加/删除DP。然后就就是了(2e, (1 + exp (e))6

6实验评价

为了评估我们的协议，我们优化了DP参数，以最小化通信。然后，我们执行微基准测试来测量所使用的每种加密方案所需的计算时间。复制我们的结果所需的代码可在 https://github.com/google-research/sparse_dp_histograms.

6.1加密方案

我们比较了AHE的两个实例：付费密码系统[29,66]和指数埃尔伽马尔密码系统[28,38]。当密文大小是一个值得关注的问题时，前者是一个众所周知的选择。后者提供更快的加密和同态操作，但解密的成本与加密值增加。这是因为x被加密为EncElGamal（g x），其中g是底层组的生成器。因此，两个密文可以通过它们相乘来同态地相加，但现在标准的埃尔伽马尔解密只产生g^{x}，所以求解x需要一个离散的对数。但是请注意，我们在x上有一个上界，因为（直到小的加性噪声项）每个桶的值将小于n。去加密所有的桶，我们可以简单地预先计算（gi）i∈[n]，并使用它来查找解密值。因此，在所有用户贡献之间平摊，在最坏的情况下，解密除了标准的ElGamal解密外，还需要一个指数幂。请注意，使用Shanks的算法[71]及其对椭圆曲线组[22,37]的优化，对于较大的n，可以进一步降低这个代价。

6.2计算和通信成本

在本节中，我们将报告我们的协议的具体计算和通信成本，以及基于混乱电路的基线解决方案。

	埃尔加马尔	佩利尔	经验。埃尔加马尔
密文大小	64字节	256字节	64字节
加密	102µs ± 1%	921µs ± 2%	152µs ± 1%
随机化（离线）	101µs ± 1%	—	—
随机化（在线）	512ns ± 1%	—	—
同音异义词添加	—	2.30µs ± 8%	537ns ± 2%
经验	101µs ± 0%	—	—
解密，解密	50.7µs ± 1%	369µs ± 1%	101µs ± 1%

表1：我们需要的加密操作的CPU微基准测试。对于指数埃尔伽马尔解密，我们报告了最坏的情况，摊销每个用户的成本。

微基准测试。在表1中，我们给出了在我们的协议中使用的加密操作的CPU时间微基准测试，以及相应的密文大小。所有实验都在英特尔Xeon铂8373CCPU@2.60 GHz，单核上运行。我们使用帕利尔和[48]的开源实现来实现我们的微基准测试[48]。为了与使用混乱电路的基线协议进行比较，我们使用EMP框架[75]在完整版本[9]的附录中实现了图14。

噪声分布和参数。回想一下，我们的方法是由每个入口噪声和重复的噪声分布参数化的。对于前者，我们使用TSDLap（•），对于后者，我们使用NBin（•）。此外，我们的协议采用了一个阈值T，以及上述分布的参数来提供隐私性（根据定理15）。

对于n个（客户端数量）和隐私参数e，6，我们对T、T、r、p，{7i}i∈[n]的值进行网格搜索，从而得到一个安全的协议实例[4]，并选择最小化整体服务器通信的配置，即最小化在步骤中插入的虚拟贡献的数量(2)在图5中。在我们的实验中，我们使用基本组合作为e来分配隐私预算（e，6）计数=eleakage如图5所示，每个步骤的=e/2（类似于6）。我们所有实验的隐私预算分割都是固定的。在本节中，我们假设是Δ = 1。

通信成本。回想一下，在我们的协议中，客户端通过一次性参与来启动协议，并且在协议执行期间不需要保持在线。我们的协议对客户端的通信成本是192字节（只要域元素适合单个密文，独立于e和n）。对于剩下的部分，我们研究服务器通信成本。在图12（左）和表2中，我们报告了任何执行我们的协议的预期总通信成本的上限。换句话说，报告的通信成本是最坏情况的输入分布，最大化通信，在实践中可能更低。例如，当在图5的步骤(4)中按伪索引进行分组时，我们假设输入中不同索引的数量为n，这就导致了该步骤的最大可能的通信。

图12（左）显示，我们的协议明显优于基于混乱电路的基线协议。事实上，基线的通信成本令人望而却步，需要超过400KB的服务器

[4]我们使用引理12的“数值”版本，在完整版本[9]的附录中称为引理18，以及在定理15中对7j的设置。

c	n	P1离线	P1在线	P1合计	P1通讯。	P2离线	P2在线	P2合计	P2通讯。	总计时间	总通信。
0.5	105	3.67	0.93	4.60	1539	0.28	0.82	1.10	141	5.70	1680
	106	1.15	0.36	1.51	482	0.26	0.26	0.51	130	2.02	612
	107	0.70	0.26	0.96	294	0.25	0.16	0.41	128	1.37	422
	108	0.56	0.23	0.78	234	0.25	0.13	0.38	128	1.16	362
	109	0.50	0.21	0.72	211	0.25	0.11	0.37	128	1.08	339
1.0	105	2.11	0.58	2.68	883	0.26	0.47	0.73	133	3.42	1016
	106	0.91	0.31	1.22	383	0.26	0.20	0.46	129	1.68	512
	107	0.63	0.24	0.87	264	0.25	0.14	0.40	128	1.27	392
	108	0.53	0.22	0.75	223	0.25	0.12	0.37	128	1.12	351
	109	0.49	0.21	0.70	206	0.25	0.11	0.36	128	1.07	334
2.0	105	1.49	0.44	1.92	624	0.26	0.33	0.59	130	2.51	754
	106	0.79	0.28	1.07	330	0.25	0.18	0.43	129	1.50	459
	107	0.59	0.23	0.82	246	0.25	0.13	0.39	129	1.21	375
	108	0.51	0.22	0.73	216	0.25	0.12	0.37	129	1.10	344
	109	0.48	0.21	0.69	203	0.25	0.11	0.36	129	1.06	332

表2：对于不同的e、n、6=值，我们的协议的每个客户端计算（以毫秒为单位）和通信成本（以字节为单位）10-11.我们使用指数元素作为加性同态加密方案。看在完整版本[9]中使用[9]。客户端成本独立于e和n，为0.46 ms的CPU时间和192字节的通信。

针对n个=的每个客户端的通信105客户端，每个客户端的成本随着客户端数量的增加而增加(设置为n个=106每个客户端需要600KB的服务器通信)。相比之下（见表2），对于e = 1，我们的协议需要大约1kb(对于n个=105)和0。5KB（用于n=106)的服务器通信。对于这些设置，这分别提高了400×和1200×。对于较大的n，我们的解决方案的每个客户端成本不断下降，导致n个≥的每个客户端成本小于0.4KB107.此外，图12（左）显示了我们的最优版本所提供的改进。特别是，对于e = 1，n =106，优化版本产生的每个客户端的总通信量少于1KB，而基本版本需要超过13KB。

图12（右）显示了对于不同的e值，总通信规模随着n是如何增长的。请注意，随着n的增长，每个客户端的成本接近于一个常数。这是0 (n)通信协议的一个很好的特性，而不是混乱的电路基线，它具有通信复杂度Ω（nlogn）（忽略对|D|的对数依赖）。更具体地说，在我们的协议中，在发送加密的客户端数据之上的开销，即图5的步骤(2)中的虚拟生成，是(n)。ø因此，随着n的增加，跨服务器的每个客户端通信接近于（恒定的）客户端通信成本。对于足够大的n，服务器所产生的成本接近所有客户端所产生的总成本。特别是，对于10亿个客户端，使用e = 1，我们的协议只需要1.07 ms的计算时间，334字节的服务器通信。

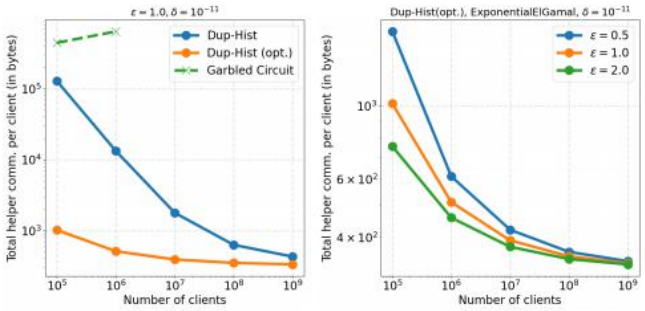

图12：我们的协议的每个客户端的成本。（左）混乱电路基线的总最坏情况通信，以及我们协议的两个变体：重复的负二项噪声（Dup-Hist），以及泊松噪声的优化变体（Dup-Hist，opt.），对于e = 1\2和6=10-11.（右）e的几个值的总服务器通信量。

7个结论和开放式问题

本文介绍了如何构造分布式双服务器协议来计算具有DP的稀疏直方图。与中央DP机制类似，我们的协议实现了通信和计算效率，这与域的大小无关，并且只与客户端贡献的数量和直方图的稀疏性成正比。通过向两个服务器显示数据的DP视图，我们的协议优于只使用混乱电路的基线。如果这是最优的，如果不是，对于任何给定的泄漏和隐私参数，其通信和计算开销的下界是什么。

虽然我们的协议可以防止客户端与两台服务器之一的任意冲突，但它并不能防止恶意客户端。在未来的工作中，防止这些问题的可能方法包括使用ZKP友好的哈希函数，切换到恶意安全的共享密钥OPRF，以及对客户端值的范围证明。

我们还提出了一个完整的正式处理的MPC协议与DP泄漏作为一个开放的问题。特别是，虽然我们的方法允许在MPC意义上组合协议，但我们仍然证明了端到端协议本身的DP，而没有组合任何子协议。具有DP泄漏的MPC协议的组合定理可以简化未来具有与我们的类似保证的协议的开发。

确认

我们感谢周明勋告诉我们本文早期版本中的一个错误。

参考文献

[1] Miklos Ajtai，贾诺斯·科姆洛斯，和恩德雷·塞梅雷迪。1983.一个O（n logn）排序网络在STOC。1－9.

[2]·弗朗西斯科·阿尔达和汉斯·乌尔里希·西蒙。2018.微分私有整数分区释放的一个下界。IPL 129（2018），1-4。

[3][3]，阿里，奥尔西尼，罗塔鲁，奈杰尔P。聪明，蒂姆木材2019.Zaphod：有效地结合LSSS和混乱电路。在WAHC。

[4]苹果和谷歌。2021.曝光通知，私人分析。https://github.com/google/exposure-notifications-android/blob/master/doc/ENPA.pdf.

[5]维克多·巴尔瑟和艾伯特·楚。2020.分离局部和洗牌差分器通过直方图的隐私。在ITC。1:1－ 1:14.

[6] Borja Balle，詹姆斯·贝尔，阿德里亚·加斯康，和科比·尼西姆。2019.隐私权洗牌模型的毯子。在加密。638－667.

[7]肯尼斯E。击球手。1968.分类网络及其应用程序。在fips春季联合计算会议。307－314.

[8] Amos贝mel，科比尼西姆和Eran Omri。2008.分布式私有数据分析：同时解决如何解决和做什么的问题。在加密。451－468.

[9]，詹姆斯·贝尔，阿德里亚·加斯康，巴迪·加齐，拉维·库马尔，帕辛·马努兰西，马里亚纳雷科娃和菲利普·肖普曼。2022.双服务器模型中的分布式的、私有的、稀疏的直方图。密码学ePrint档案（2022年）。https://eprint.iacr.org/2022/920

[10]，詹姆斯·亨利·贝尔，卡利斯塔·A。博纳维茨，阿德里亚加斯康，坦克克雷德角，和马-伊安娜·雷科娃。2020.使用（多边形）对数开销保护单服务器聚合。在中国化学会。

[11]安德里亚·比塔乌，乌尔法尔·埃林松，马尼亚提斯，伊利亚·米罗诺夫，阿南斯·拉古-内森，大卫李，米奇鲁道米纳，乌沙斯里科德，朱利安廷恩斯，和伯恩哈德西菲尔德。2017.普罗希洛：在人群中拥有分析人员的强大隐私权。在SOSP。441－459.

[12]耶利米·布洛克基，阿努帕姆·达塔，和约瑟夫·邦诺。2016.不同的私有密码频率列表。在NDSS。

[13]丹·博格达诺夫，马尔科·乔梅特，桑德·西伊姆和梅里尔·瓦赫特。2016.利用真实的数据量在云中进行隐私保护的税务欺诈检测。网络论研究，报告，https://cyber.ee/research/reports/T-4-24-Privacy-preserving-tax-fraud-detection-in-the-cloud-with-realisticdata-volumes.pdf.

[14] Jonas·博勒和弗洛里安·克施鲍姆。2021.安全的多方计算不同的私人重打手。在中国化学会。2361－2377.

[15]基思·博纳维茨，弗拉基米尔·伊万诺夫，本·克鲁特，安东尼奥·马塞东尼，H。布伦丹麦克马汉，萨瓦尔帕特尔，丹尼尔拉马奇，亚伦西格尔，和卡恩赛斯。2017.实用的安全聚合，为保护隐私的机器学习。在中国化学会。

[16]丹·Boneh，埃莱特·博伊尔，亨利·科里根-吉布斯，尼夫·吉尔博亚和尤瓦尔·伊沙伊。2021.轻重量级技术的私人重量级打击者。在SP。762－776.

[17] Elette Boyle，尼夫·吉尔博亚，和Yuval Ishai。2016.功能秘密共享：改进-设备和扩展。在中国化学会。1292－ 1303.

[18]伦纳特布劳恩，丹尼尔德姆勒，托马斯施耐德，奥莱克桑德和特卡琴科。2022.运动--一个混合协议多方计算的框架。ACM跨。Priv。秒25，2（2022），1－35.

[19] Mark Bun，Kobbi尼西姆和斯特默。2019.同时进行私人学习的多个概念。JMLR 20（2019），94:1－94:34.

[20]·贝内迪克特·邦兹，乔纳森·布特尔，丹·博内，安德鲁·波尔斯特拉，彼得·维尔，和格雷格麦克斯韦。2018.防报：机密交易的简短证明和更多。在SP。

[21] TH..－陈慧慧，施惠莲和黎明之歌。2012.最佳下界不同的私人多党聚集。在欧空局。277－288.

[22]，查奇安尼斯，康斯坦丁诺斯和查尔基亚斯和瓦莱里亚·尼古拉尼科。2021.通过压缩的离散日志查找表在块链中进行同态解密。在DPM/CBT@经典。328－339.

[23] Albert Cheu，亚当·D。史密斯，乔纳森R。乌尔曼，大卫·泽伯和马克西姆日利亚夫。2019.通过洗牌来实现的分布式差异性隐私。在欧洲地下室。375－403.

[24] Albert Cheu和马克西姆·日里亚耶夫。2021.不同的私有直方图在来自伪造用户的洗牌模型。CoRR abs/2104.02739（2021）。

[25]格雷厄姆胸衣，塞西莉亚普罗科皮克，斯里瓦斯塔瓦和桑。L.Tran。2012.对于稀疏数据的差异私有摘要。在ICDT。

[26]亨利·科里根-吉布斯和丹·博纳。2017.Prio：私有的、健壮的和可扩展的总统计数据的计算。在NSDI。

[27]亨利科里根-吉布斯，丹博内，加里陈，史蒂文恩格尔哈特，罗伯特赫尔默，克里斯·胡顿-查普斯基，宫口安东尼，埃里克·雷斯科拉，和彼得·圣安德烈。2020.保护隐私的火狐遥测与Prio。https：//rwc。iacr.org/2020/slides/Gibbs.pdf.

[28]罗纳德·克莱默，罗萨里奥·热纳罗，和贝瑞·舍恩的制造商。1997.安全和最佳高效的多权力选举方案。在欧洲墓穴，卷。1233.103－ 118.

[29]，伊万·丹加德和Mads Jurik。2001.一个概括，简化和帕利尔概率公钥系统的一些应用。在PKC，卷。1992.119－ 136.

[30]达米恩达斯，詹姆斯沃斯，布莱恩特吉普森，和钦莫伊曼达扬。2022.不同的私有分区选择。PoPETS 2022，1（2022），339－352.

[31]辛西娅德工作，克里什纳拉姆肯塔帕迪，弗兰克麦克雪利，伊利亚米罗诺夫，和莫尼Naor。2006.我们的数据，我们自己：通过分布式噪声产生的隐私。在欧洲地下室。486－503.

[32]辛西娅Dwork，弗兰克麦克雪利，科比尼西姆，和亚当史密。2006.Calibrat-私有数据分析中的噪声与敏感性。在TCC。

[33]乌尔法尔·埃林森，维塔利·费尔德曼，伊利亚·米罗诺夫，阿纳斯·拉古纳坦，库纳尔塔尔瓦尔和塔库尔塔。2019.洗牌放大：从本地到中心的差异隐私。在SODA。2468－2479.

[34]乌尔法尔·埃林森，瓦西尔·皮胡尔和亚历山德拉·科罗洛娃。2014.RAPPOR：运行域名化的可聚合的隐私-保护的顺序响应。在中国化学会。

[35]亚历山大·埃夫菲米耶夫斯基，约翰内斯·格尔克和罗摩克里什南·斯里坎特。2003.限制隐私保护数据挖掘中的隐私侵犯。在PODS。211－222.

[36] Vitaly费尔德曼，库纳尔和奥纳尔·麦克米兰。2021.隐藏在克隆：一个简单而近乎最优的洗牌隐私放大分析。在FOCS。954－964.

[37]史蒂文D。加尔布雷斯、王平、张方国。2017.利用改进的婴儿巨步算法计算椭圆曲线离散对数。Adv。数学通勤。11，3（2017），453－469.

[38] Taher El Gamal。1984.一种基于公钥密码系统和签名方案离散对数。在密码卷。196.10－ 18.

[39] Badih Ghazi，诺亚·戈洛维奇，拉维·库马尔，帕辛·马努兰西，拉斯马斯·帕格，和Ameya维林克。2020.来自匿名信息的纯差异的私人总结。在ITC。15:1－ 15:23.

[40] Badih Ghazi，诺亚·戈洛维奇，拉维·库马尔，拉斯马斯·帕格，和阿米亚·维林克。2021.论多重匿名信息的力量：差异隐私模糊模型中的频率估计与选择。在欧洲地下室。463－488.

[41] Badih Ghazi，本克鲁特，拉维库马尔，马努兰西，彭嘉宇，叶夫根尼斯克沃尔佐夫，姚王，和赖特。2022.多方到达和频率直方图：私有、安全和实用。PoPETS 2022，1（2022），373－395.

[42] Badih Ghazi，拉维·库马尔，和帕辛·马努兰西。2021.用户级别差异通过相关抽样进行私人学习。在神经IPS。

[43] Badih Ghazi，拉维·库马尔，帕辛·马努兰西，和拉斯马斯·帕格。2020.私有的从匿名消息中计数：接近最优的准确性与消失的通信开销的准确性。在ICML。3505－3514.

[44]过时的戈德里奇。2006.密码学的基础：第1卷。剑桥美国大学出版社。

[45]过时的戈德里奇。2009.密码学的基础：第2卷，基本应用。剑桥大学出版社

[46]采访了戈尔德里奇、西尔维奥·米卡利和阿维·维格德森。1987.如何玩任何精神游戏游戏在STOC。218－229.

[47]，戈德瓦瑟，米卡利和拉克夫。1985.知识的复杂性交互式证明系统。在STOC。

[48]谷歌。2019.私人加入和计算。https://github.com/google/private-joinand-compute/.

[49]洛伦佐·格拉斯，德米特里·科夫拉托维奇，克里斯蒂安·雷克伯格，阿纳布·罗伊，和马库斯·肖夫内格。2021.波塞冬：零知识证明系统的一个新的哈希函数。在USENIX。519－535.

[50]亚当·格罗斯，彼得·林达尔和迈克·罗苏莱克。2019.更便宜的私人设置部分通过差异私人泄漏。PoPETS 2019，3（2019），6－25.

[51]迈克尔海，李超，米克劳，大卫D。延森2009.准确专用网络的度分布的估计。在ICDM。169－ 178.

[52]迈克尔·海，维布尔·拉斯托吉，杰罗姆·米克劳和丹·苏丘。2010.Boosting通过一致性确定差异私有直方图的准确性。VLDB 3，1（2010），1021－ 1032.

[53]严黄，大卫·埃文斯和乔纳森·卡茨。2012.专用设置交叉点：混乱的电路比定制的协议更好吗？。在NDSS。

[54][54]和刘晓敏。2010.集合交叉口的快速安全计算。在硫氰酸盐。

[55]湿婆，K。李，科比尼西姆，索菲亚拉斯科德尼科娃和亚当·史密斯。2008.我们可以私下学习什么？。在FOCS。

[56]马塞尔凯勒。2020.MP-SPDZ：一个多功能的多方计算框架。在中国化学会。

[57]亚历山德拉科罗洛娃，克里什纳拉姆肯塔帕迪，尼娜米什拉，和亚历山大ntoula。2009.私下发布搜索查询和单击。在WWW。

[58]瓦西里奥斯兰波斯，安德鲁C米勒，史蒂夫克罗桑，和克里斯蒂安斯蒂芬森。2015.使用搜索查询日志在临近预测流感样疾病发病率方面的进展。科学报告12760（2015）。问题5。

[59] Yehuda林德尔。2017.如何模拟它——个教程上的模拟证明技术。在关于密码学基础的教程中。施普林格国际出版公司，277-346。

[60] Yehuda林德尔。2020.安全的多方计算。CACM 64，1（2020），86－96.[61] Pasin马努兰西。2022.对不同的私人匿名化的严格界限直方图在SOSA。203－213.

[62] Sahar Mazloom和S。多夫戈登。2018.使用有差异的方式进行安全计算私人访问模式。在中国化学会。490－507.

[63]弗兰克·麦克雪利和拉图尔·马哈詹。2010.差异性-专用网络跟踪
分析。Sigmomput。通勤。发动机的旋转40，4（aug 2010），123- 134.
[64]凯瑟琳。草地1986.一个更有效的密码匹配
在没有持续可用的情况下使用的协议。在SP。134- 137.
[65]伊利亚·米罗诺夫，奥姆坎特·潘迪，奥默·雷林戈尔德和萨利尔·瓦德汉。
2009.计算微分的隐私。在加密。
[66]帕斯卡·佩利尔。1999.基于复合学位的公钥密码系统
残差等级。在欧洲墓穴，卷。1592.223–238.
[67]、杨宁、李宁辉。2013.了解层次结构
不同私有直方图的cal方法。VLDB 6，14（2013）.
[68]伊多罗斯，丹尼尔·诺布尔，布雷特海门威福克和安德烈亚斯海伯伦。2019.
蜜脆：没有受信任的核心的大规模差异私有聚合。在SOSP。
[69]江户罗斯，张恒楚，安德烈亚斯·海伯伦和本杰明·C。刺穿2020.
果园：规模的私人分析。在OSDI。
[70]菲利普肖普曼，伦纳特沃格尔桑，阿德里亚加斯康，和博尔贾巴尔。2020.
分布式数据库上的安全和可扩展的文档相似性：拯救的差异隐私。PoPETS 2020，
2（2020），209–229.
[71]丹尼尔香克斯。1971.类数，一个因子分解的理论，和属。在程序中。
Symp。数学Soc。1971年，卷。20.41–440.
[72]AnandaTheersh。2019.不同的私人匿名直方图。
在神经IPS。7969–7979.
[73] Sameer Wagh，习近平，阿什文，和米塔尔。2021.dp密码学：在新兴应用中结
合差异隐私和密码学。CACM 64，2（2021），84–93.
王天浩、耶利米、李宁辉、周浩。2017.局部的
频率估计的差异私有协议。在USENIX。
[75]小王，Alex J。马洛泽莫夫和乔纳森·卡茨。2016.高效的多方计算工具包。
https://github.com/emp-toolkit .
[76]小王，塞缪尔·拉内鲁奇和乔纳森·卡茨。2017.全球规模的安全
多方计算。在中国化学会。
[77]AndrewChihYao。1986.如何产生和交换秘密。在FOCS。
162- 167.
张军，小奎，谢兴。2016.一个不同的私有的
层次结构分解的算法。在SIGMOD。155- 170.
朱[79]文南，彼得·凯鲁兹，布伦丹·麦克马汉、孙海成和李伟。
2020.联邦重杀手发现与不同的隐私。在测试中。3837–3847.