

Distributed Private Heavy Hitters

Justin Hsu*

Sanjeev Khanna†

Aaron Roth‡

May 9, 2012

Abstract

In this paper, we give efficient algorithms and lower bounds for solving the *heavy hitters* problem while preserving *differential privacy* in the fully distributed *local* model. In this model, there are n parties, each of which possesses a single element from a universe of size N . The heavy hitters problem is to find the identity of the most common element shared amongst the n parties. In the local model, there is no trusted database administrator, and so the algorithm must interact with each of the n parties separately, using a differentially private protocol. We give tight information-theoretic upper and lower bounds on the accuracy to which this problem can be solved in the local model (giving a separation between the local model and the more common centralized model of privacy), as well as computationally efficient algorithms even in the case where the data universe N may be exponentially large.

1 Introduction

Consider the problem of a website administrator who wishes to know what his most common traffic sources are. Each of n visitors arrives with a single *referring site*: the name of the last website that she visited, which is drawn from a vast universe N of possible referring sites (N here is the set of all websites on the internet). There is value in identifying the most popular referring site (the *heavy hitter*): the site administrator may be able to better tailor the content of his webpage, or better focus his marketing resources. On the other hand, the identity of each individual's referring site might be embarrassing or otherwise revealing, and is therefore private information. We can therefore imagine a world in which this information must be treated “privately.” Moreover, in this situation, visitors are communicating directly with the servers of the websites that they visit: i.e. there is no third party who might be trusted to aggregate all of the referring website data and provide privacy preserving statistics to the website administrator. In this setting, how well can the website administrator estimate the heavy hitter while being able to provide formal privacy guarantees to his visitors?

This situation can more generally be modeled as the *heavy hitters* problem under the constraint of *differential privacy*. There are n individuals $i \in [n]$ each of whom is associated with an element $v_i \in N$ of some large data universe N . The *heavy hitter* is the most frequently occurring element $x \in N$ among the set $\{v_1, \dots, v_n\}$, and we would like to be able to identify that element, or one

*Department of Computer and Information Sciences, University of Pennsylvania. Email: justhsu@cis.upenn.edu

†Department of Computer and Information Sciences, University of Pennsylvania. Email: sanjeev@cis.upenn.edu

‡Department of Computer and Information Sciences, University of Pennsylvania. Email: aaroth@cis.upenn.edu

Supported in part by NSF Awards CCF-1101389 and CNS-1065060.

that occurs almost as frequently as the heavy hitter. Moreover, we wish to solve this problem while preserving *differential privacy* in the fully distributed (local) model. We define this formally in section 2, but roughly speaking, an algorithm is differentially private if changes to the data of single individuals only result in small changes in the output distribution of the algorithm. Moreover, in the fully distributed setting, each individual (who can be viewed as a database of size 1) must interact with the algorithm independently of all of the other individuals, using a differentially private algorithm. This is in contrast to the more commonly studied centralized model, in which a trusted database administrator may have (exact) access to all of the data, and coordinate a private computation.

We study this problem both from an information theoretic point of view, and from the point of view of efficient algorithms. We say that an algorithm for the private heavy hitters problem is *efficient* if it runs in time $\text{poly}(n, \log N)$: i.e. polynomial in the database size, but only polylogarithmic in the universe size (i.e. in what we view as the most interesting range of parameters, the universe may be exponentially larger than the size of the database). We give tight information theoretic upper and lower bounds on the accuracy to which the heavy hitter can be found in the private distributed setting (separating this model from the private centralized setting), and give several efficient algorithms which achieve good, although information-theoretically sub-optimal accuracy guarantees. We leave open the question of whether *efficient* algorithms can exactly match the information theoretic bounds we prove for the private heavy hitters problem in the distributed setting.

1.1 Our Results

In this section, we summarize our results. The bounds we discuss here are informal and hide many of the parameters which we have not yet defined. The formal bounds are given in the main body of the paper.

First, we provide an information theoretic characterization of the accuracy to which any algorithm (independent of computational constraints) can solve the heavy hitters problem in the private distributed setting. We say that an algorithm is α -accurate if it returns a universe element which occurs with frequency at most an additive α smaller than the true heavy hitter. In the centralized setting, a simple application of the exponential mechanism [MT07] gives an α -accurate mechanism for the heavy-hitters problem where $\alpha = O(\log |N|)$, which in particular, is independent of the number of individuals n . In contrast, we show that in the fully distributed setting, no algorithm can be α -accurate for $\alpha = \Omega(\sqrt{n})$ even in the case in which $|N| = 2$. Conversely, we give an almost matching upper bound (and an algorithm with run-time linear in N) which is α -accurate for $\alpha = O(\sqrt{n \log N})$.

Next, we consider *efficient* algorithms which run in time only polylogarithmic in the universe size $|N|$. Here, we give two algorithms. One is an application of a compressed sensing algorithm of Gilbert et al [GLPS10], which is α -accurate for $\alpha = \tilde{O}(n^{5/6} \log \log N)$. Then, we give an algorithm based on group-testing using pairwise independent hash functions, which has an incomparable bound. Roughly speaking, it guarantees to return the exact heavy hitter (i.e. $\alpha = 0$) whenever the frequency of the heavy hitter is larger than the ℓ_2 -norm of the frequencies of the remaining elements. Depending on how these frequencies are distributed, this can correspond to a bound of α -accuracy for α ranging anywhere between the optimal $\alpha = O(\sqrt{n})$ to $\alpha = O(n)$.

1.2 Our Techniques

Our upper bounds, both information theoretic, and those with efficient algorithms, are based on the general technique of *random projection* and *concentration of measure*. To prove our information theoretic upper bound, we observe that to find the heavy hitter, we may view the private database as a histogram v in N dimensional space. Then, it is enough to find the index $i \in [N]$ of the universe element which maximizes $\langle v, e_i \rangle$, where e_i is the i 'th standard basis vector. Both v and each e_i have small ℓ_1 -norm, and so each of these inner products can be approximately preserved by taking a random projection into $\tilde{O}(\log N)$ dimensional space. Moreover, we can project each individual's data into this space independently in the fully distributed setting, incurring a loss of only $O(\sqrt{n})$ in accuracy. This mechanism, however, is not efficient, because to find the heavy hitter, we must enumerate through all $|N|$ basis vectors e_i in order to find the one that maximizes the inner product with the projected database. Similar ideas lead to our efficient algorithms, albeit with worse accuracy guarantees. For example, in our first algorithm, we apply techniques from compressed sensing to the projected database to recover (approximately) the heavy hitter, rather than checking basis vectors directly. In our second algorithm, we take a projection using a particular family of pairwise-independent hash functions, which are linear functions of the data universe elements. Because of this linearity, we are able to efficiently “invert” the projection matrix in order to find the heavy hitter.

Our lower bound separates the distributed setting from the centralized setting by applying an anti-concentration argument. Roughly speaking, we observe that in the fully distributed setting, if individual data elements were selected uniformly *i.i.d.* from the data universe N , then even after conditioning on the messages exchanged with any differentially private algorithm, they remain independently distributed, and approximately uniform. Therefore, by the Berry-Esseen theorem, even after any algorithm computes its estimate of the heavy hitter, the true distribution over counts remains approximately normally distributed. Since the Gaussian distribution exhibits strong anti-concentration properties, this allows us to unconditionally give an $\Omega(\sqrt{n})$ lower bound for any algorithm in the fully distributed setting.

1.3 Related Work

Differential privacy was introduced in a sequence of papers culminating in [DMNS06], and has since become the standard “solution concept” for privacy in the theoretical computer science literature. There is by now a very large literature on this topic, which is too large to summarize here. Instead, we focus only on the most closely related work, and refer the curious reader to a survey of Dwork [Dwo08].

Most of the literature on differential privacy focuses on the *centralized* model, in which there is a trusted database administrator. In this paper, we focus on the *local* or *fully distributed* model, introduced by [KLN⁺08] and studied also by [BNO08], in which each individual holds their own data (i.e. there are n databases, each of size 1), and the algorithm must interact with each one in a differentially private manner. There has been little work in this more restrictive model—the problems of *learning* [KLN⁺08] and *query release* [GHRU11] in the local model are well understood¹, but only

¹Roughly, the set of concepts that can be *learned* in the local model given polynomial sample complexity is equal to the set of concepts that can be learned in the SQ model given polynomial query complexity [KLN⁺08], and the set of queries that can be *released* in the local model given polynomial sample complexity is equal to the set of concepts that can be agnostically learned in the SQ model given polynomial query complexity [GHRU11], but the polynomials

up to polynomial factors that do not imply tight bounds for the heavy hitters problem. The *two-party* setting (which is intermediate between the centralized and fully distributed setting), in which the data is divided between two databases without a trusted central administrator, was considered by [MMP⁺10]. They proved a separation between the two-party setting and the centralized setting for the problem of computing the Hamming distance between two strings. In this work, we prove a separation between the fully distributed setting and the centralized setting for the problem of estimating the heavy hitter.

A variant of the private heavy hitters problem has been considered in the setting of *pan-private streaming algorithms* [DNP⁺10, MMNW11]. This work considers a different (although related) problem in a different (although related) setting. [DNP⁺10, MMNW11] consider a setting in which a stream of elements is presented to the algorithm, and the algorithm must estimate the *approximate count* of frequently occurring elements (i.e. the number of “heavy hitters”). In this setting, the universe elements themselves are the individuals appearing in the stream, and so it is not possible to reveal the identity of the heavy hitter. In contrast, in our work, individuals are distinct from universe elements, which merely label the individuals. Moreover, our goal here is to actually identify a specific universe element which is the heavy hitter, or which occurs almost as frequently. Also, [DNP⁺10, MMNW11] work in the centralized setting, but demand *pan-privacy*, which roughly requires that the internal state of the algorithm itself remain differentially private. In contrast, we work in the *local privacy* setting which gives a guarantee which is strictly stronger than pan-privacy. Because algorithms in the local privacy setting only interact with individuals in a differentially private way, and never have any other access to the private data, any algorithm in the local privacy model can never have its state depend on data in a non-private way, and such algorithms therefore also preserve pan-privacy. Therefore, our upper bounds hold also in the setting of pan-privacy, whereas our lower bounds do not necessarily apply to algorithms which only satisfy the weaker guarantee of pan-privacy.

Finally, we note that many of the upper bound techniques we employ have been previously put to use in the centralized model of data privacy i.e. random projections [BLR08, BR11] and compressed sensing (both for lower bounds [DMT07] and algorithms [LZwy11]). As algorithmic techniques, these are rarely optimal in the centralized privacy setting. We remark that they are particularly well suited to the fully distributed setting which we study here, because in a formal sense, algorithms in the local model of privacy are constrained to only access the private data using noisy linear queries, which is exactly the form of access used by random linear projections and compressed sensing measurements.

2 Preliminaries

A database v consists of n records from a data universe N , one corresponding to each of n individuals: for $i \in [n]$, $v^i \in N$ and $v = \{v^1, \dots, v^n\}$ which may be a *multiset*. Without loss of generality, we will index the elements of the data universe from 1 to $|N|$. It will be convenient for us to represent databases as *histograms*. In this representation, $v \in \mathbb{N}^{|N|}$, where v_i represents the number of occurrences of the i 'th universe element in the database. Further, we write $v^i \in \mathbb{N}^{|N|}$ for each individual $i \in [n]$, where $v_j^i = 1$ if individual i is associated with the j 'th universe element, and $v_{j'}^i = 0$ for all other $j' \neq j$. Note that in this histogram notation, we have: $v = \sum_{i=1}^n v^i$. In

are not equal.

the following, we will usually use the histogram notation for mathematical convenience, with the understanding that we can in fact more concisely represent the database as a multiset.

Given a database v , the *heavy hitter* is the universe element that occurs most frequently in the database: $hh(v) = \arg \max_{i \in N} v_i$. We refer to the frequency with which the heavy hitter occurs as $fhh(v) = v_{hh(v)}$. We want to design algorithms which return universe elements that occur *almost as frequently as the heavy hitter*.

Definition 2.1. An algorithm A is (α, β) -accurate for the heavy hitters problem if for every database $v \in \mathbb{N}^{|N|}$, with probability at least $1 - \beta$: $A(v) = i^*$ such that $v_{i^*} \geq fhh(v) - \alpha$.

2.1 Differential Privacy

Differential privacy constrains the sensitivity of a randomized algorithm to individual changes in its input.

Definition 2.2. An algorithm $A : \mathbb{N}^{|N|} \rightarrow R$ is (ϵ, δ) -differentially private if for all $v, v' \in \mathbb{N}^{|N|}$ such that $\|v - v'\|_1 \leq 1$, and for all events $S \subseteq R$:

$$\Pr[A(v) \in S] \leq \exp(\epsilon) \Pr[A(v') \in S] + \delta$$

Typically, we will want δ to be negligibly small, whereas we think of ϵ as being a small constant (and never smaller than $\epsilon = O(1/n)$).

A useful distribution is the *Laplace* distribution.

Definition 2.3 (The Laplace Distribution). The Laplace Distribution (centered at 0) with scale b is the distribution with probability density function $\text{Lap}(x|b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right)$. We will sometimes write $\text{Lap}(b)$ to denote the Laplace distribution with scale b , and will sometimes abuse notation and write $\text{Lap}(b)$ simply to denote a random variable $X \sim \text{Lap}(b)$.

A fundamental result in data privacy is that perturbing low sensitivity queries with Laplace noise preserves $(\epsilon, 0)$ -differential privacy.

Theorem 2.4 ([DMNS06]). *Suppose $Q : \mathbb{N}^{|N|} \rightarrow \mathbb{R}$ is a function such that for all databases $v, v' \in \mathbb{N}^{|N|}$ such that $\|v - v'\|_1 \leq 1$, $|Q(v) - Q(v')| \leq c$. Then the procedure which on input v releases $Q(v) + X$, where X is a draw from a $\text{Lap}(c/\epsilon)$ distribution, preserves $(\epsilon, 0)$ -differential privacy.*

It will be useful to understand how privacy parameters for individual steps of an algorithm compose into privacy guarantees for the entire algorithm. The following useful theorem is a special case of a theorem proven by Dwork, Rothblum, and Vadhan:

Theorem 2.5 (Privacy Composition [DRV10]). *Let $0 < \epsilon, \delta < 1$, and let M_1, \dots, M_T be $(\epsilon', 0)$ -differentially private algorithms for some $\epsilon' \leq \epsilon / \sqrt{8T \log(\frac{1}{\delta})}$. Then the algorithm M which on input v outputs $M(v) = (M_1(v), \dots, M_T(v))$ is (ϵ, δ) -differentially private.*

The local privacy model (alternately, the fully distributed setting) was introduced by Kairouz et al. [KLN⁺08] in the context of learning. The local privacy model formalizes randomized response: there is no central database of private data. Instead, each individual i maintains possession of their own data element (i.e. a database v^i of size $\|v^i\|_1 = 1$), and answers questions about it only in a differentially private manner. Formally, the database $v \in \mathbb{N}^{|N|}$ is the sum of n databases of size 1: $v = \sum_{i=1}^n v^i$, and each v^i is held by individual i .

Definition 2.6 ([KLN⁺08] (Local Randomizer)). An (ϵ, δ) -local randomizer $R : \mathbb{N}^{|N|} \rightarrow R$ is an (ϵ, δ) -differentially private algorithm that takes a database of size $\|v\|_1 = 1$.

In the local privacy model, algorithms may interact with the database only through a local randomizer oracle:

Definition 2.7 ([KLN⁺08] (LR Oracle)). An LR oracle $LR_v(\cdot, \cdot)$ takes as input an index $i \in [n]$ and an (ϵ, δ) -local randomizer R and outputs a random value $w \in R$ chosen according to the distribution $R(v^i)$, where v^i is the element held by the i 'th individual in the database.

Definition 2.8 ([KLN⁺08] (Local Algorithm)). An algorithm is (ϵ, δ) -local if it accesses the database v via the oracle LR_v , that satisfies the following restriction: if $LR_v(i, R_1), \dots, LR_v(i, R_k)$ are the algorithm's invocations of LR_v on index i , then the joint outputs of each of these k algorithms must be (ϵ, δ) -differentially private.

To avoid cumbersome notation, we will avoid the formalism of LR oracles, instead remembering that for algorithms in the local model, any operation on v^i must be carried out without access to any v^j for $j \neq i$, and must be differentially private in isolation.

2.2 Probabilistic Tools

We will make use of several useful probabilistic tools. First, the well-known Johnson-Lindenstrauss lemma:

Theorem 2.9 (Johnson-Lindenstrauss Lemma). *Let $0 < \gamma < 1$ be given. For any set V of q vectors in \mathbb{R}^N , there exists a linear map $A : \mathbb{R}^N \rightarrow \mathbb{R}^m$ with $m = O\left(\frac{\log q}{\gamma^2}\right)$ such that A is approximately an isometric embedding of V into \mathbb{R}^m . That is, for all $x, y \in V$, we have the two bounds*

$$(1 - \gamma)\|x - y\|^2 \leq \|A(x - y)\|^2 \leq (1 + \gamma)\|x - y\|^2$$

$$|\langle Ax, Ay \rangle - \langle x, y \rangle| \leq O(\gamma(\|x\|^2 + \|y\|^2))$$

In particular, any $m \times N$ random projection matrix A_p , whose entries are drawn IID uniformly from $\{-1/\sqrt{m}, 1/\sqrt{m}\}$, enjoys this property with probability at least $1 - \beta$, with $m = O\left(\frac{\log q \log(1/\beta)}{\gamma^2}\right)$. Note that this projection matrix does not depend on the set of vectors V .

In other words, any set of q points in a high dimensional space can be *obliviously* embedded into a space of dimension $O(\log q)$ such that w.h.p. this embedding essentially preserves pairwise distances.

In our analysis, we will also make use of a simple tail bound on the sums of Laplace random variables:

Theorem 2.10 (See, e.g. [GRU12]). *Let $X_i, i \in [n]$ be IID random variables drawn from the $\text{Lap}(b)$ (the Laplace distribution with parameter b) and let $X = \sum_{i=1}^n X_i$. Then, we have the bound*

$$\Pr[X \geq T] \leq \begin{cases} \exp\left(-\frac{T^2}{6nb^2}\right) & : T \leq nb \\ \exp\left(-\frac{T}{6b}\right) & : T > nb \end{cases}$$

In particular, choosing $T_\beta = b\sqrt{6n} \log(2/\beta)$ gives

$$\Pr[|X| \leq T_\beta] \geq 1 - \beta$$

3 Information Theoretic Upper and Lower Bounds.

In this section we present upper and lower bounds on the accuracy to which any algorithm in the fully distributed model can privately approximate heavy hitters. Our upper bound can be viewed as an algorithm, albeit one that runs in time linear in $|N|$ and so is not what we consider to be efficient.

3.1 An Upper Bound via Johnson-Lindenstrauss Projections

We present here our first algorithm, referred to as *JL-HH*, that solves the heavy hitters problem in the local model using the Johnson-Lindenstrauss lemma. The algorithm JL-HH is outlined in Algorithm 1. We write e_i to refer to the i 'th standard basis vector in \mathbb{R}^N , and write $\text{RandomProjection}(m, N + 1)$ for a subroutine which returns a linear embedding of $N + 1$ points into m dimensions using a random $\pm 1/\sqrt{m}$ valued projection matrix, as specified by the Johnson-Lindenstrauss lemma. By the Johnson-Lindenstrauss lemma, for any set of $N + 1$ elements, this map approximately preserves pairwise distances with high probability.

Algorithm 1 JL-HH Mechanism

Input: Private histograms $v^i \in \mathbb{N}^N, i \in [n]$. Privacy parameters $\epsilon, \delta > 0$. Failure probability $\beta > 0$.

Output: p^* , index of the heavy hitter.

```

 $\gamma \leftarrow 1/n^2$ 
 $m \leftarrow \frac{\log(N+1) \log(2/\beta)}{\gamma^2}$ 
 $A \leftarrow \text{RandomProjection}(m, N + 1)$ 
for  $p = 1$  to  $N$  indices do
    for  $i = 1$  to  $n$  users do
         $z^i \sim \left\{ \text{Lap} \left( \frac{\sqrt{8 \log(1/\delta)}}{\epsilon} \right) \right\}^m$ 
         $q^i = Av^i + z^i$ 
         $r_{ip} = \langle Ae_p, q^i \rangle$ 
    end for
     $c_p \leftarrow \sum_{i=1}^n r_{ip}$ 
end for
 $p^* \leftarrow \text{argmax}_p c_p$ 
return  $p^*$ 
```

JL-HH is based on the following straightforward idea. If v is a private histogram, we will estimate the count of the i 'th element ($\langle v, e_i \rangle$), by estimating $\langle Av, Ae_i \rangle$, and returning the largest count. By Theorem 2.9, since we are using the random projections matrix, we have that with high probability, inner products between points in the set $V = \{e_1 \cdots e_N, v\}$ are approximately preserved under A . However, we cannot access Av directly since v is private data. To preserve differential privacy, our mechanism must add noise z to Av , and work only with the noisy samples. Our analysis will thus focus on bounding the error introduced by this noise term. First, though, we show that JL-HH is differentially private.

Lemma 3.1. *JL-HH operates in the local privacy model and is (ϵ, δ) -differentially private.*

Proof. The measurement Av is computed in the fully distributed setting, by computing $Av \approx \sum_{i=1}^n Av^i + z^i$. Each individual i may compute $Av^i + z^i$ which corresponds to answering a sequence of m linear queries, each with sensitivity $1/\sqrt{m}$. By Theorem 2.4, the noise that JL-HH adds guarantees that each such query is ϵ_0 -differentially private, with

$$\epsilon_0 = \frac{\epsilon}{\sqrt{8m \log(1/\delta)}}$$

Thus, by Theorem 2.5, this composition is (ϵ, δ) -differentially private, as desired. From here, the algorithm works with the noised measurement instead of private data, and is therefore differentially private. \square

Now, we show that JL-HH estimates the counts to within an additive error of $O\left(\frac{\sqrt{n \log N}}{\epsilon}\right)$.

Theorem 3.2. *For any $\beta > 0$, JL-HH mechanism is (α, β) -accurate for the heavy hitters problem, with $\alpha = O\left(\frac{\sqrt{n \log(N/\beta) \log(1/\delta)}}{\epsilon}\right)$.*

Proof. Let v be the private histogram, and let $z = \sum_{i=1}^n z^i$ denote the sum of the noise vectors added to each individual's data v^i . The error of the mechanism is at most

$$2 \max_{i \in [N]} |\langle e_i, v \rangle - \langle Ae_i, Av + z \rangle|$$

Note that for all j , the random variable z_j is distributed as the sum of n i.i.d. Laplace random variables each with scale $b = \sqrt{8 \log 1/\delta}/\epsilon$. To calculate the error for an index i , we may write:

$$|\langle e_i, v \rangle - \langle Ae_i, Av + z \rangle| \leq |\langle e_i, v \rangle - \langle Ae_i, Av \rangle| + |\langle Ae_i, z \rangle| \quad (1)$$

$$= O(\gamma \|v\|^2 + |\langle Ae_i, z \rangle|) \quad (2)$$

with the second equality following from Theorem 2.9. Recall that we have set $\gamma = n^{-2}$, and let A be the random projection matrix, with $m = O(\log N \log(2/\beta)/\gamma^2)$. With probability at least $1 - \beta/2$, the random projections matrix A actually satisfies the property for the Johnson-Lindenstrauss lemma. So, we have

$$\langle Ae_i, z \rangle = \sum_{j=1}^m (Ae_i)_j \sum_{i=1}^n z_j^i$$

But Ae_i is a vector of length m with entries drawn uniformly from $\pm 1/\sqrt{m}$. Since the Laplace distribution is also symmetric, the distribution of this sum is identical to the distribution of a sum of mn i.i.d. Laplace random variables each with scale $b = \frac{\sqrt{8 \log 1/\delta}}{\sqrt{m\epsilon}}$. By our tail bound in Theorem 2.10, with probability at least $1 - \beta/2N$, this sum is bounded by $O\left(\frac{\sqrt{n \log(1/\delta) \log(N/\beta)}}{\epsilon}\right)$.

On the other hand, the other error $|\langle e_i, v \rangle - \langle Ae_i, Av \rangle|$ can be bounded by Equation (2), and hence is $O(1)$ by our choice of γ . Thus, with probability at least $1 - \beta/2N$, we have that the estimated count for index i is within an additive factor of $O\left(\frac{\sqrt{n \log(1/\delta) \log(N/\beta)}}{\epsilon}\right)$ to the true count of index i . Taking a union bound over all indices, we have that with probability at least $1 - \beta/2$, this accuracy holds for the heavy hitter, and all other elements. Since the probability of failing when picking A was at most $\beta/2$, this gives the desired high probability bound. \square

It is worthwhile to compare JL-HH with a more naive approaches. A simpler differentially private algorithm to solve the distributed heavy hitters problem is to have each user simply add noise $\text{Lap}(1/\epsilon)$ to each entry in the user's private histogram, and report this vector to the central party, which sums the noisy vectors and estimates the most frequently occurring item. This is differentially private, as any neighboring histogram will change exactly one entry in a user's histogram. However, this method requires having each user transmit $O(N)$ amount of information to the central party. JL-HH achieves similar accuracy compared to this naive approach, but since the clients compress the histogram first, only $O(\log N)$ information must be communicated. Even though JL-HH runs in time linear in N , there are natural situations where long running time can be tolerated, but large communication complexity cannot, for instance, if the central party is a server farm with considerable computational resources, but the communication with users must happen over standard network links.

3.2 A Lower Bound via Anti-Concentration

Here we show that our upper bound in the previous subsection is essentially optimal: for any $\epsilon < 1/2$ and any δ bounded away from 1 by a constant, no (ϵ, δ) -private mechanism in the fully distributed setting can be α -accurate for the heavy hitters problem for some $\alpha = \Omega(\sqrt{n})$, even in the case in which $|N| = 2$. Our theorem follows by arguing that even after conditioning on the output of the differentially private interaction with each individual in the local model, there is still quite a bit of uncertainty in the distribution over heavy hitters, if the universe elements were initially distributed uniformly at random. We take advantage of this uncertainty to apply an anti-concentration argument, which implies that no matter what answer the algorithm predicts, there is enough randomness leftover in the database instance that the algorithm is likely to be incorrect (with at least some constant probability β). We remark that our technique (while specific to the local privacy model) holds for (ϵ, δ) -differential privacy, even when $\delta > 0$. This is similar to lower bounds based on reconstruction arguments [DN03], and in contrast to other techniques for proving lower bounds in the centralized model, such as the elegant packing arguments used in [BKN10, HT10], which are specific to $(\epsilon, 0)$ -differential privacy. We use an independence argument also used by [MMP⁺10] to prove a lower bound in the two-party setting, and by [BNO08] to prove a lower bound in the fully distributed setting.

Theorem 3.3. *For any $\epsilon \leq 1/2$ and $\delta < 1$ bounded away from 1, there exists an $\alpha = \Omega(\sqrt{n})$ and a $\beta = \Omega(1)$ such that no (ϵ, δ) -private mechanism in the local model is (α, β) -accurate for the heavy hitters problem.*

Proof. We give a lower bound instance in which the universe is $N = \{0, 1\}$. Each individual i is assigned a universe element $s_i \in \{0, 1\}$ uniformly at random. Let $A_i : N \rightarrow \mathcal{M}$ denote the (ϵ, δ) -differentially private algorithm which acts on the data s_i of individual i , and write $m_i = A_i(s_i)$.

We condition on the order of the parties that we query and on the output of each algorithm, $m_i = \hat{m}_i$ for fixed $\hat{m}_i \in \mathcal{M}$.

We first observe that conditioning on the outputs of each A_i : $m_i = \hat{m}_i$ for each i , the random variables s_i remain independent of one another. (This is a standard fact from communication complexity)

We next argue that under this conditioning, the marginal distributions of a constant fraction of the s_i variables remain approximately uniform. If we define the random variables X_i to be the

indicator of the event $s_i = \hat{s}_i$ (conditioning on all the messages), we can apply Bayes' rule to get for all $i \in [n]$:

$$\begin{aligned}\Pr[X_i = \hat{s}_i] &= \Pr[s_i = \hat{s}_i | m_i = \hat{m}_i] \\ &= \frac{\Pr[m_i = \hat{m}_i | s_i = \hat{s}_i] \Pr[s_i = \hat{s}_i]}{\Pr[m_i = \hat{m}_i]} \\ &\leq \frac{\Pr[m_i = \hat{m}_i | s_i = \hat{s}_i] \Pr[s_i = \hat{s}_i]}{\Pr[m_i = \hat{m}_i | s_i = b]}\end{aligned}$$

where b is some element of the universe. Because each A_i is (ϵ, δ) -differentially private, we have that with probability at least $1 - \delta$, the following random variable (where the randomness is over the choice of \hat{m}_i) is bounded:

$$\frac{\Pr[m_i = \hat{m}_i | s_i = \hat{s}_i]}{\Pr[m_i = \hat{m}_i | s_i = b]} \leq e^\epsilon$$

and thus with probability $1 - \delta$ over the choice of \hat{m}_j : $\Pr[X_i = \hat{s}_i] \leq (e^\epsilon)/2$, using the prior on s_i .

In similar fashion, we can prove a lower bound on the probability. So, we have that for each i independently with probability at least $1 - \delta$: $\Pr[X_i = \hat{s}_i] \in [(e^{-\epsilon})/2, (e^\epsilon)/2]$. Because we assume $\epsilon \leq 1/2$, we therefore have for each i independently with probability $1 - \delta$: $\Pr[X_i = \hat{s}_i] \in [c_1, c_2]$ where c_1, c_2 are constants bounded away from 0 and 1 respectively. Because this occurs with constant $1 - \delta$ probability for each i , for any constant β , we can (by the Chernoff bound) take n to be sufficiently large so that except with probability $\beta/2$, we have $\Pr[X_i = \hat{s}_i] \in [c_1, c_2]$ for $\Omega(n)$ individuals i . This, together with the conditional independence of the X_i 's, allows us to apply the Berry-Esseen theorem:

Theorem 3.4 (Berry-Esseen). *Given independent random variables $X_i, i \in [n]$, let $\mu_i = \mathbb{E}[X_i]$, $\sigma_i^2 = \mathbb{E}[(X_i - \mu_i)^2]$, $\beta_i = \mathbb{E}[|X_i - \mu_i|^3]$, and let*

$$S_n = \frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$$

If F_n is the cdf of S_n , and Φ is the cdf for the standard normal distribution, then there exists a constant C such that

$$\sup_x |F_n(x) - \Phi(x)| \leq C\psi$$

where

$$\psi = \left(\sum_{i=1}^n \sigma_i^2 \right)^{-1/2} \max \frac{\beta_i}{\sigma_i^2}$$

For each of the $\Omega(n)$ individuals i for which $\Pr[X_i = 1] \in [c_1, c_2]$, each σ_i^2 and β_i is a constant bounded away from 0. Thus, we have with probability at least $\beta/2$: $\psi \leq O(1/\sqrt{n})$, and hence the cdf F_n of the sample mean S_n converges uniformly to the normal distribution. By a change of variables, this means that the cdf of the sum $\sum_{i=1}^n (X_i - \mu_i)$ converges to the cdf of a normal distribution with mean 0 and variance $\sigma^2 = \sum_{i=1}^n \sigma_i^2 = \Omega(n)$. The next lemma lower bounds the probability that S_n is within an additive factor of $\Omega(\sqrt{n})$ of its mean.

Lemma 3.5. Let $\beta > 0$ be given and condition on the event that $\Pr[X_i = 1] \in [c_1, c_2]$ for $\Omega(n)$ individuals $i \in [n]$. For sufficiently large n , there exists a constant C such that

$$\Pr \left[\left| \sum_{i=1}^n (X_i - \mu_i) \right| \geq C\sqrt{n} \right] \geq 1 - \beta/2$$

of Lemma. This is immediate, since by the Berry-Esseen theorem the sum $\sum_{i=1}^n (X_i - \mu_i)$ converges uniformly to a Gaussian distribution with standard deviation $\sigma = \Omega(\sqrt{n})$. \square

To complete the proof, we note that the distribution of $n_1 \equiv \sum_{i=1}^n X_i$ is simply the distribution of the number of occurrences of universe element 1, after conditioning on the outcome of differentially private mechanisms A_1, \dots, A_n . Consider a mechanism, which given the outcome of mechanisms A_1, \dots, A_n attempts to guess the value of n_1 , and outputs \hat{n}_1 . Let $\mu = \sum_{i=1}^n \mu_i$. By the properties of the Gaussian distribution we have:

$$\Pr[|n_1 - \hat{n}_1| \leq t] \leq \Pr[|n_1 - \mu| \leq t]$$

for all values of t . In particular, for some $t = C\sqrt{n}$ we have shown that this probability is at most β . In other words, we have shown that for some constant $\beta \geq 0$ and for some $\alpha = \Omega(\sqrt{n})$, there is no (ϵ, δ) -private algorithm in the local model which is able to estimate the *frequency* of the heavy hitter to within an additive α factor with probability $1 - \beta$. It is straightforward to see that there therefore cannot be an (α, β) -accurate, (ϵ, δ) -private mechanism for the heavy hitters problem: any such mechanism could be converted to a mechanism which estimates the frequency of the heavy hitter by introducing “dummy” individuals corresponding to the universe element which is not the heavy hitter, and performing a binary search over their count by computing the identity of the heavy hitter in each dummy instance. The count at which the identity of the heavy hitter in the dummy instance changes can then be used to estimate the frequency of the true heavy hitter. \square

4 Efficient Algorithms

In the last section, we saw the Johnson-Lindenstrauss algorithm which gave almost optimal accuracy guarantees, but had running time linear in $|N|$. In this section, we consider efficient algorithms with running time $\text{poly}(n, \log |N|)$. The first is an application of a sublinear time algorithm from the compressed sensing literature, and the second is a group-testing approach made efficient by the use of a particular family of pairwise-independent hash functions.

4.1 GLPS Sparse Recovery

In this section we adapt a sophisticated algorithm from compressed sensing. Gilbert, et al. [GLPS10] present a sparse recovery algorithm (we refer to it as the GLPS algorithm) that takes linear measurements from a sparse vector, and reconstructs the original vector to high accuracy. Importantly, the algorithm runs in time polylogarithmic in $|N|$, and polynomial in the sparsity parameter of the vector. We remark that our database v is n -sparse: it has at most n non-zero components. In the rest of this section, we will write v_s to denote the vector v truncated to contain only its s largest components.

Let s be a sparsity parameter, and let γ be a tunable approximation level. The GLPS algorithm runs in time $O((s/\gamma) \log^c N)$, and makes $m = O(s \log(N/s)/\gamma)$ measurements from a specially constructed (randomized) $\{-1, 0, 1\}$ valued matrix, which we will denote Φ . Given measurements $u = \Phi v + z$ (where z is arbitrary noise), the algorithm returns an approximation \hat{v} , with error guarantee

$$\|v - \hat{v}\|_2 \leq (1 + \gamma)\|v - v_s\|_2 + \gamma \log(s) \frac{\|z\|_2}{\kappa} \quad (3)$$

with probability at least $3/4$, where $\kappa = O(\log^2(s) \log(N/s))$. Though the GLPS bound only occurs with probability $3/4$, the success probability can be made arbitrarily close to 1 by running this algorithm several times. In particular, using the amplification lemma from [GLM⁺10], the failure probability can be driven down to β at a cost of only a factor of $\log(1/\beta)$ in the accuracy. In what follows, we analyze a single run of the algorithm, with $\gamma = 1$.

Algorithm 2 GLPS-HH Mechanism

Input: Private histograms $v^i \in \mathbb{N}^N, i \in [n]$. GLPS matrix Φ . Privacy parameters $\epsilon, \delta > 0$.

Output: p^* , estimated index of heavy hitter.

```

 $m \leftarrow s \log(N/s)$ 
 $b \leftarrow \sqrt{8m \log(1/\delta)} / \epsilon$ 
for  $i = 1$  to  $n$  users do
     $z^i \sim \{\text{Lap}(b)\}^m$ 
     $q^i \leftarrow \Phi v^i + z^i$ 
end for
 $c \leftarrow \sum_{i=1}^n q^i$ 
 $\hat{v} \leftarrow \text{GLPS}(c, \Phi)$ 
 $p^* \leftarrow \text{argmax}_p \hat{v}_p$ 
return  $p^*$ 
```

Next, we will show that GLPS-HH is (ϵ, δ) -differentially private.

Theorem 4.1. *GLPS-HH operates in the local privacy model and is (ϵ, δ) -differentially private.*

Proof. The algorithm operates in the local privacy model because each individual i compute $\Phi v^i + z^i$ independently, which corresponds to answering m linear queries, each with sensitivity 1. The magnitude of the Laplace noise added, z^i , is then sufficient (by Theorem 2.5) to guarantee (ϵ, δ) -differential privacy for each individual. \square

Next, we will bound the error that we introduce by adding noise for differential privacy.

Theorem 4.2. *Let $\beta > 0$ be given. GLPS-HH is $(\alpha, 3/4 - \beta)$ -accurate for the heavy hitters problem, with*

$$\alpha = O\left(\frac{n^{5/6} \log^{1/3}(1/\beta) \log \log N \log^{1/6}(1/\delta)}{\epsilon^{1/3}}\right)$$

Proof. Let $b = \sqrt{8m \log(1/\delta)} / \epsilon$. Let v denote the combined private database, and let \hat{v} denote the estimated private database returned by GLPS. GLPS-HH uses the GLPS algorithm with measurements $c = \Phi v + z, z = \sum_i z^i$, where the noise vector z has each entry drawn from $\sum_{i=1}^n \text{Lap}(b)$. From Theorem 2.10, we have the bound (for a fixed index i)

$$\Pr[|z_i| \leq O(b\sqrt{n} \log(m/\beta))] \geq 1 - \beta/m$$

Taking a union bound over all m indices, we find this bound holds over all components with probability at least $1 - \beta$. Thus we can bound

$$\|z\|_2 \leq O(b\sqrt{nm} \log(m/\beta)) = O\left(\frac{s \log(N/s) \log(s \log(N/s)/\beta) \sqrt{n \log(1/\delta)}}{\epsilon}\right)$$

With probability $3/4$, we have the GLPS bound Equation (3) with $\gamma = 1$, from which we can estimate

$$\|v - \hat{v}\|_\infty \leq \|v - \hat{v}\|_2 \leq 2\|v - v_s\|_2 + O\left(\frac{s \log(s \log(N/s)/\beta) \sqrt{n \log(1/\delta)}}{\epsilon \log s}\right)$$

By a Lemma from [GSTV07], we have $\|v - v_s\|_2 \leq \|v\|_1 / \sqrt{s}$. Now, in the worst case, $\|v\|_1 = O(n)$, and we need to choose s to balance the errors in

$$\|v - \hat{v}\|_\infty \leq 2\frac{n}{\sqrt{s}} + O\left(\frac{s \log(s \log(N/s)/\beta) \sqrt{n \log(1/\delta)}}{\epsilon \log s}\right)$$

By setting s to be:

$$s = \left(\frac{\epsilon}{\log(1/\beta)} \sqrt{\frac{n}{\log(1/\delta)}}\right)^{2/3}$$

when we get an error bound

$$\|v - \hat{v}\|_\infty \leq O\left(\frac{n^{5/6} \log^{1/3}(1/\beta) \log \log N \log^{1/6}(1/\delta)}{\epsilon^{1/3}}\right)$$

Thus, with probability at least $3/4 - \beta$, we get the desired accuracy. \square

4.2 The Bucket mechanism

In this section we present a second computationally efficient algorithm, based on group-testing and a specific family of pairwise independent hash functions.

At a high level, our algorithm, referred to as the *Bucket mechanism*, runs $O(\log(1/\beta))$ trials consisting of $O(\log N)$ 0/1 valued hash functions in each trial. For a given trial, the mechanism hashes each universe element into one of two buckets for each hash function. Then, the mechanism tries to find an element that hashes into the bucket with more weight (the *majority bucket*) for all the hash functions. If there is such an element, it is a candidate for the heavy hitter for that trial. Finally, the mechanism takes a majority vote over the candidates from each trial to output a final heavy hitter.

For efficiency purposes we do not use truly random hash functions, but instead rely on a particular family of pairwise-independent hash functions which can be expressed as linear functions on the bits of a universe element. Specifically, each function h in the family maps $[N]$ to $\{0, 1\}$, and is parameterized by a bit-string $r \in \{0, 1\}^{\log |N|}$. In particular, given any bit-string $r \in \{0, 1\}^{\log |N|}$, we define $h_r(x) = \langle r, b(x) \rangle$, where $b(x)$ denotes the binary representation of x . r is chosen uniformly

Algorithm 3 The Bucket Mechanism

Input: Private labels $v^i \in [N], i \in [n]$. Failure probability $\beta > 0$. Privacy parameters $\epsilon, \delta > 0$.

Output: p^* , the index of the heavy hitter.

```
 $F \leftarrow \{0, 1\}^{\log N} \setminus 0$ 
for  $i = 1$  to  $8 \log(1/\beta)$  trials do
   $H \in \{0, 1\}^{\log(12N) \times \log N} \leftarrow$  Draw  $\log(12N)$  rows from  $F$ , uniformly at random.
   $u \in \mathbb{R}^{\log(12N)} \leftarrow 0$ 
  for  $j = 1$  to  $n$  users do
     $b \in \{0, 1\}^{\log N} \leftarrow$  binary expansion of  $v^j$ .
     $s \leftarrow Hb \pmod{2}$ 
     $z \sim \left\{ \text{Lap} \left( \frac{8\sqrt{\log(12N) \log(1/\beta) \log(1/\delta)}}{\epsilon} \right) \right\}^{\log(12N)}$ 
     $u \leftarrow u + s + z$ 
  end for
  for  $k = 1$  to  $\log(12N)$  hash functions do
     $b_k \leftarrow \begin{cases} 1 & : u_k > n/2 \\ 0 & : \text{otherwise} \end{cases}$ 
  end for
   $w_i \leftarrow \begin{cases} x_0 & : Hx_0 = b \pmod{2} \\ \perp & : Hx = b \pmod{2} \text{ infeasible} \end{cases}$ 
  end for
   $w^* \leftarrow$  most frequent  $w_i$ , ignoring  $\perp$ 
return  $p^* \leftarrow w^*$  converted from binary
```

at random from the set of all strings $r \in \{0, 1\}^{\log |N|} \setminus 0^{\log |N|}$. Given hash functions of this form, and a list of target buckets, the problem of finding an element that hashing to all of the target buckets is equivalent to solving a linear system mod 2, which can be done efficiently. Our family of hash functions operates on the element label in binary, hence the conversions to and from binary in the algorithm.

We will now show that the bucket mechanism is (ϵ, δ) -differentially private, runs in time $\text{poly}(n, \log |N|)$, and assuming a certain condition on the distribution over universe elements, returns the exact heavy hitter. The accuracy analysis proceeds in two steps: first, we argue that with constant probability $> 1/2$, the heavy hitter is the unique element hashed into the larger bucket by every hash function in a given trial. Then, we argue that with high probability, the proceeding event indeed occurs in the majority of trials, and so the majority vote among all trials returns the true heavy hitter.

Theorem 4.3. *The Bucket mechanism operates in the local model and is (ϵ, δ) -differentially private.*

Proof. Each party answers $\log(12N)$ 1-sensitive queries about only their own data for each trial, with a total of $8\log(1/\beta)$ trials. By Theorem 2.5, the correct amount of noise is added to preserve (ϵ, δ) -differential privacy. \square

Theorem 4.4. *For fixed $\epsilon, \delta > 0$ and failure probability $\beta > 0$, the Bucket mechanism runs in time $O(n \log(1/\beta) \log^3 N)$.*

Proof. The step that dominates the run time is the inner loop over each party. For each user, the algorithm evaluates $O(\log N)$ hash functions. Each evaluation calculates the inner product of two $\log N$ -length bit strings, and there are $O(\log N)$ hash functions. So, each user takes time $\log^2 N$ per trial. With n users and $O(\log(1/\beta))$ trials, the result follows. \square

We first prove a simple tail bound on sums of k -wise independent random variables, modifying a result given by Bellare and Rompel, [BR94].

Lemma 4.5. *Let k be even. Take a k -independent set of random variables X_i , with $0 \leq X_i \leq c_i$, let $X = \sum X_i$, and let $\mu = \mathbb{E}[X]$. We have:*

$$\Pr[|X - \mu| > t] \leq C_k \left(\frac{ck}{t^2} \right)^{k/2}$$

with $c = \sum c_i^2$, and $C_k = 2\sqrt{\pi k}e^{k/2-1/(6k)} \leq 1.0004$.

Proof. By Markov's inequality, we can write:

$$\Pr[|X - \mu| > t] = \Pr[(X - \mu)^k > t^k] \leq \frac{\mathbb{E}[(X - \mu)^k]}{t^k}$$

However, if we expand out the product, we find that we only need to consider the expected value of products of at most k of the variables X_i . Thus, without loss of generality, we may consider X_i to be independent for the following calculation.

$$\begin{aligned}
\mathbb{E}[(X - \mu)^k] &= \int_0^\infty \Pr[(X - \mu)^k > s] ds \\
&= \int_0^\infty \Pr[|X - \mu| > s^{1/k}] ds \\
&\leq \int_0^\infty 2 \exp\left(-\frac{s^{2/k}}{2 \sum c_i^2}\right) ds
\end{aligned}$$

where we have used that the X_i are independent in order to apply Azuma's inequality. By a change of variables, and letting $c = \sum c_i^2$, we have

$$\begin{aligned}
\mathbb{E}[(X - \mu)^k] &\leq \int_0^\infty k(2c)^{k/2} e^{-y} y^{k/2-1} dy \\
&= (2c)^{k/2} k \Gamma(k/2 - 1) \\
&= 2(2c)^{k/2} \left(\frac{k}{2}\right)! \\
&\leq 2c^{k/2} \sqrt{\pi k} \left(\frac{k}{e}\right)^{k/2} e^{1/6k}
\end{aligned}$$

where we have used Stirling's approximation in the last step. Now, we get

$$\Pr[|X - \mu| > t] = \frac{\mathbb{E}[(X - \mu)^k]}{t^k} \leq C_k \left(\frac{ck}{t^2}\right)^{k/2}$$

as desired. \square

Lemma 4.6. *Let $\beta, \epsilon, \delta > 0$ be given, and consider a single trial in the Bucket mechanism. Without loss of generality, suppose that the elements are labeled in decreasing order of count, with counts $v_1 \geq v_2 \geq \dots \geq v_N$. Write $c = \sum_{i=2}^N v_i^2$, let k_1 be the number of hash functions per trial, and k_2 be the number of trials. If we have the condition*

$$v_1 \geq 2 \sqrt{\frac{12k_1 c}{\beta}} + b(k_1, k_2) \sqrt{6n} \log\left(\frac{6k_1}{\beta}\right)$$

where b is the parameter for (ϵ, δ) -differential privacy:

$$b(k_1, k_2) = \frac{\sqrt{8k_1 k_2 \log(1/\delta)}}{\epsilon}$$

then with probability at least $1 - 2\beta/3$, the heavy hitter is hashed into the larger bucket for each hash function in the trial.

Proof. First consider a single hash function. If we define random variables $X_i, i \in [N]$ by:

$$X_i = \begin{cases} v_i & : i \text{ is hashed to bucket 1} \\ 0 & : \text{otherwise} \end{cases}$$

and the function $f(X_2, \dots, X_N) = \sum_{i=2}^N X_i$, we show that the true heavy hitter will be hashed to the larger bucket (with high probability) if f does not deviate from the mean by too much. If f is close to the mean, then no matter which bucket the heavy hitter is hashed to, that will become the larger bucket. However, we will need to keep track of the noise that will be added to preserve differential privacy. We want v_1 to be large enough to overcome the noise (with high probability).

More precisely, by Theorem 2.10, the sum of n Laplace noise terms will be bounded by $b\sqrt{6n}\log(6k_1/\beta)$, with probability at least $1 - \beta/3k_1$. We also know that the collection $\{X_2, \dots, X_N\}$ is a pairwise-independent set of random variables, so applying Lemma 4.5 with $X = f$, and $t = \sqrt{\frac{12k_1 c}{\beta}}$, we have that

$$\Pr[|f - \mu| > t] \leq C_2 \left(\frac{2c}{t^2} \right) \leq 4 \left(\frac{c}{t^2} \right) = \frac{\beta}{3k_1}$$

with C_2 a constant from Lemma 4.5. The difference between the counts in the two buckets will be $2|f - \mu|$, so for the heavy hitter to be hashed to the larger bucket, we need $v_1 \geq 2|f - \mu| + |z|$, where z is the Laplace noise term, with high probability. Taking a union bound over k_1 hash functions, we have that

$$2|f - \mu| + |z| \leq 2\sqrt{\frac{12k_1 c}{\beta}} + b\sqrt{6n}\log\left(\frac{6k_1}{\beta}\right)$$

holds for all the hash functions in this trial with probability at least $1 - 2\beta/3$. But by assumption, v_1 is larger than this gap, and so we are done. \square

Lemma 4.7. *Let the notation be as in the previous Lemma, and consider a single trial in the Bucket mechanism. If we set $k_1 = \log\left(\frac{3N}{\beta}\right)$, then with probability at least $1 - \beta/3$, no other element will be hashed to the same bucket as the heavy hitter through all the hash functions.*

Proof. Pick any element g besides the heavy hitter, and consider a single hash function. Since the hash function is pairwise-independent, conditioning on where the heavy hitter is hashed will not change the marginal for where g will be hashed. Thus, there is a $1/2$ chance of g colliding with the heavy hitter for any given hash function. Since the hash functions are drawn independently at random, the chance of this collision happening on every function is $(1/2)^{k_1} = \beta/(3N)$, by choice of k_1 . Taking a union bound over the $N - 1$ elements besides the heavy hitter, we have that this collision probability for all elements is bounded by $\beta/3$, as desired. \square

Now, we are ready to put everything together.

Theorem 4.8. *Let the notation be as in the previous Lemma. If we set $k_1 = \log(12N)$, $k_2 = 8\log(1/\beta)$, and if we have the condition*

$$v_1 \geq 8\sqrt{2c\log(12N)} + \frac{8\log(24\log(12N))\sqrt{6n\log(12N)\log(1/\beta)\log(1/\delta)}}{\epsilon} = \\ \tilde{\Omega}\left(\frac{\sqrt{\log|N|}\left(\sqrt{c} + \sqrt{n\log\frac{1}{\beta}\log\frac{1}{\delta}}\right)}{\epsilon}\right)$$

then the Bucket mechanism is $(0, \beta)$ -accurate for the heavy hitters problem.

Proof. First, note that k_1 and the condition have been chosen so that from Lemmas 4.6 and 4.7, for any single trial, the heavy hitter is always hashed to the larger bucket, and is the unique such element, with probability at least $3/4$. These two conditions ensure that we are able to correctly identify the heavy hitter with probability $3/4$ for a single trial. Now, as the trials are independent, we apply a Chernoff bound to show that out of k_2 Bernoulli variables with success probability $3/4$, the probability that at least half of them succeed is bounded below by

$$\Pr[\text{Majority Vote Success}] \geq 1 - e^{-2k_2(1/4)^2} = 1 - \beta$$

by our choice of k_2 . Thus, the Bucket mechanism returns the true heavy hitter with probability at least $1 - \beta$. \square

We note that the accuracy guarantee of the bucket mechanism is incomparable to those of our other mechanisms. While the other mechanisms guarantee (without conditions) to return an element which occurs within some additive factor α as frequently as the true heavy hitter, the bucket mechanism always returns the true heavy hitter, so long as a certain condition on v is satisfied. When the condition is not satisfied, the algorithm comes with no guarantees. The condition is roughly that the heavy hitter should occur more frequently than the ℓ_2 -norm of the frequencies of all other elements. Depending on the distribution over elements, this condition can be satisfied when the heavy hitter occurs with frequency as small as $\tilde{O}(\sqrt{n})$, or can require frequency as large as $\Omega(n)$. Finally, we note that this condition is not unreasonable. It will, for example, be satisfied with high probability if the frequency of the database elements follows a power law distribution, such as a Zipf distribution.

5 Discussion and Open Questions

We have initiated the study of the *private heavy hitters* problem in the fully distributed (local) privacy model. We have provided an (almost) tight characterization of the accuracy to which the problem can in principle be solved. In particular, we have separated the local privacy model from the centralized privacy model: we have shown that even the easier problem of simply releasing the approximate count of the heavy hitter cannot be accomplished to accuracy better than $\Omega(\sqrt{n})$ in the local model, whereas this can be accomplished to $O(1)$ accuracy in the centralized model. We have also given several efficient algorithms for the heavy hitters problem, but these algorithms do not in general achieve the tight $\tilde{O}(\sqrt{n \log |N|})$ accuracy bound that we have established is possible in principle. We leave open the question of whether there exist *efficient algorithms* in the local model which can solve the heavy hitters problem up to this information theoretically optimal bound.

Acknowledgments

We would like to thank Martin Strauss for providing valuable clarifications and insights about [GSTV07], and the anonymous reviewers for their helpful suggestions. We would also like to thank Andreas Haeberlen for suggesting that we study the heavy hitters problem in the fully distributed setting, and Andreas, Marco Gaboardi, Benjamin Pierce, and Arjun Narayan for valuable discussions.

References

- [BKN10] A. Beimel, S. Kasiviswanathan, and K. Nissim. Bounds on the sample complexity for private learning and private data release. *Theory of Cryptography*, pages 437–454, 2010.
- [BLR08] A. Blum, K. Ligett, and A. Roth. A learning theory approach to non-interactive database privacy. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 609–618. ACM, 2008.
- [BNO08] A. Beimel, K. Nissim, and E. Omri. Distributed private data analysis: Simultaneously solving how and what. *Advances in Cryptology—CRYPTO 2008*, pages 451–468, 2008.
- [BR94] M. Bellare and J. Rompel. Randomness-efficient oblivious sampling. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, pages 276–287. IEEE Computer Society, 1994.
- [BR11] A. Blum and A. Roth. Fast private data release algorithms for sparse queries. *CoRR*, abs/1111.6842, 2011.
- [DMNS06] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Theory of Cryptography Conference TCC*, volume 3876 of *Lecture Notes in Computer Science*, page 265. Springer, 2006.
- [DMT07] C. Dwork, F. McSherry, and K. Talwar. The price of privacy and the limits of LP decoding. In *Proceedings of the thirty-ninth annual ACM Symposium on Theory of Computing*, page 94. ACM, 2007.
- [DN03] I. Dinur and K. Nissim. Revealing information while preserving privacy. In *22nd ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS)*, pages 202–210, 2003.
- [DNP⁺10] C. Dwork, M. Naor, T. Pitassi, G.N. Rothblum, and S. Yekhanin. Pan-private streaming algorithms. In *In Proceedings of ICS*, 2010.
- [DRV10] C. Dwork, G.N. Rothblum, and S. Vadhan. Boosting and differential privacy. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science*, pages 51–60. IEEE, 2010.
- [Dwo08] C. Dwork. Differential Privacy: A Survey of Results. In *Theory and Applications of Models of Computation—TAMC 2008*, volume 4978 of *Lecture Notes In Computer Science*, pages 1–19, 2008.
- [GHRU11] A. Gupta, M. Hardt, A. Roth, and J. Ullman. Privately Releasing Conjunctions and the Statistical Query Barrier. In *Proceedings of the 43rd annual ACM Symposium on the Theory of Computing*. ACM New York, NY, USA, 2011.
- [GLM⁺10] A. Gupta, K. Ligett, F. McSherry, A. Roth, and K. Talwar. Differentially Private Combinatorial Optimization. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 2010.

- [GLPS10] A.C. Gilbert, Y. Li, E. Porat, and M.J. Strauss. Approximate sparse recovery: optimizing time and measurements. In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 475–484. ACM, 2010.
- [GRU12] A. Gupta, A. Roth, and J. Ullman. Iterative constructions and private data release. In *Proceedings of the Ninth IACR Theory of Cryptography Conference (TCC)*, 2012.
- [GSTV07] A.C. Gilbert, M.J. Strauss, J.A. Tropp, and R. Vershynin. One sketch for all: fast algorithms for compressed sensing. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 237–246. ACM, 2007.
- [HT10] M. Hardt and K. Talwar. On the Geometry of Differential Privacy. In *The 42nd ACM Symposium on the Theory of Computing, 2010. STOC’10*, 2010.
- [KLN⁺08] S.P. Kasiviswanathan, H.K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What Can We Learn Privately? In *IEEE 49th Annual IEEE Symposium on Foundations of Computer Science, 2008. FOCS’08*, pages 531–540, 2008.
- [LZwy11] Y.D. Li, Z. Zhang, M. Winslett, and Y. Yang. Compressive mechanism: Utilizing sparse representation in differential privacy. In *Proceedings of the 10th annual ACM workshop on Privacy in the electronic society*, pages 177–182. ACM, 2011.
- [MMNW11] D. Mir, S. Muthukrishnan, A. Nikolov, and R.N. Wright. Pan-private algorithms via statistics on sketches. In *Proceedings of the 30th symposium on Principles of database systems of data*, pages 37–48. ACM, 2011.
- [MMP⁺10] A. McGregor, I. Mironov, T. Pitassi, O. Reingold, K. Talwar, and S. Vadhan. The limits of two-party differential privacy. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 81–90. IEEE, 2010.
- [MT07] F. McSherry and K. Talwar. Mechanism design via differential privacy. In *Proceedings of the 48th Annual Symposium on Foundations of Computer Science*, 2007.