Toward Privacy in Public Databases
(Draft)

Shuchi Chawla
Carnegie Mellon University
shuchi@cs.cmu.edu

Cynthia Dwork
Microsoft Research SVC
dwork@microsoft.com

Frank McSherry
Microsoft Research SVC
mcsherry@microsoft.com

Adam Smith
Massachusetts Institute of Technology
asmith@theory.lcs.mit.edu

Hoeteck Wee
University of California, Berkeley
hoeteck@cs.berkeley.edu

May 31, 2004

Abstract

We begin a theoretical study of the *census problem*. Informally, in a census individual respondents give private information to a trusted party (the census bureau), who publishes a *sanitized* version of the data. There are two fundamentally conflicting requirements: *privacy* for the respondents and *utility* of the sanitized data. Unlike in the study of secure function evaluation, in which privacy is preserved to the extent possible given a specific functionality goal, in the census problem *privacy* is paramount; intuitively, things that cannot be learned "safely" should not be learned at all.

An important contribution of this work is a definition of privacy (and privacy compromise) for statistical databases, together with a method for describing and comparing the privacy offered by specific sanitization techniques. We obtain several privacy results using two different sanitization techniques, and then show how to combine them via cross training. We also obtain two utility results involving clustering.

1 Introduction

We begin a theoretical study of the *census problem*. Informally, in a census individual respondents give private information to a trusted party (the *census bureau*), who publishes a *sanitized* version of the data. There are two fundamentally conflicting requirements: *privacy* for the respondents and *utility* of the sanitized data. While both require formal definition, their essential tension is clear: perfect privacy can be achieved by publishing nothing at all – but this has no utility; perfect utility can be obtained by publishing the data exactly as received from the respondents, but this offers no privacy. Very roughly, the sanitization should permit the data analyst to identify strong stereotypes, while preserving the privacy of individuals.

This is not a new problem. Database privacy was heavily studied by computer scientists in the seventies and eighties (see [2] for an excellent survey). One of the classical techniques, query monitoring, was recently reconsidered by [31]. An entire issue of the Journal of Official Statistics was dedicated to this subject in 1998. The cryptography community has a longstanding engagement in the study of related topics such as private information retrieval and secure function evaluation. The database community began investigating privacy-preserving data mining in the year 2000. We find much of this work unsatisfying:

- Much of the statistics literature, as well as the beautiful algorithmic work of Gusfield [28], is concerned with identifying and protecting sensitive entries in contingency tables [21, 6, 7, 8, 9, 15, 16, 17, 34]. In the 2-dimensional case, these are $m \times n$ tables of statistics, together with the row and column sums (called *marginals*). Certain entries are identified as sensitive, and the data in the sensitive entries need to be hidden, either by suppression or alteration (see [2] or [11] for a more complete treatment). As an example, suppose we have a database containing a large number of records, with many different attributes per record. For a given pair of attribute types, one taking, say, three possible values and the other taking four possible values, the contingency table would be a 3×4 matrix giving the number of records in the database satisfying each of the 12 pairs of attribute values, together with the marginals. It is common to regard an entry as sensitive if it corresponds to a pair of attribute values that has support at most a fixed number, typically 1 or 2. Because of the common desire to publish exact row and column sums, to protect sensitive cells it is frequently necessary to suppress some non-sensitive cells as well, and much work has been devoted to the question of finding a minimal set of entries to suppress so as to protect the sensitive entries. In a related vein, there is also a vast literature on local move iterative simulation methods in which a random process is used to generate new tables with identical contingency tables [13, 15]. The reason for regarding low-count cells as sensitive is to prevent linkage with other databases: if a given pair of attribute values uniquely identifies an individual, then these fields can be used as a key in other databases to retrieve further information about the individual. This reasoning makes sense to us. However, we are unaware of any clear definition of privacy or theory of when entries should be regarded as sensitive. We have also not seen a clear specification of the information available to the adversary whose goal is to learn the values of sensitive cells.

- The literature also contains a number of output perturbation techniques. In this setting, a query control mechanism receives queries, computes exact answers, and then outputs a perturbed answer as the response to the query [12, 5]. Although generally dismissed as insecure, intuitively, because noise added in response to multiple queries can cancel out (see [14] for quantitative results; see also Section 5 of [2]), the subject is experiencing a revival. Roughly speaking, insecurity arises when the number of queries may be large as a function of the size of the database. See [14] and [19] for rigorous positive results in the case that the adversary is restricted to a sublinear number of queries – a reasonable assumption when databases are enormous.

- Query monitoring – monitoring queries to ensure that, even in the context of previous queries, they do not reveal sensitive information – is computationally intractable [31] and may fail to protect privacy, for example, in the setting in which the adversary knows even one real database record [14].

- In k-anonymity, data are sanitized by suppressing fields from individual records (replacing their values with a special symbol) and aggregating categories (e.g., instead of having two categories, Ages 6-10 and 11-15, there is just one category, Ages 6-15), so that the following condition holds: every row in the sanitized database (such a row may contain special symbols) appears at least k times in the sanitized database [35]. The intuition is that k-anonymity provides privacy protection by guaranteeing that each released record will relate to at least k individuals, even if the records are directly linked to external information. While appealing, this notion is problematic. For example, based on which attributes have been suppressed and the underlying distribution on data, it may be possible to guess the suppressed attributes of a record (for example, if the suppressed attribute is sex and the number of identical records with sex suppressed is two).

- As pointed out in [20], perhaps the very first database paper on privacy-preserving data mining ([4]) failed to take into account the knowledge an adversary may have about the database, such as, for example, the fact that ages are positive.
- In secure function evaluation, privacy is preserved only to the extent possible given a specific functionality goal; but which functions are "safe" in the context of statistical databases? The literature is silent on this question.

In Section 6 we discuss three papers in the literature, all written concurrently with the research effort reported here, that are consonant with our point of view (that is, we like the definitions). As we will discuss, the definitions in those papers, while similar to each other, differ significantly from ours.

A Common Suggestion. We assume a *real database* (RDB) of n samples, each chosen independently according to a fixed underlying distribution $\mathcal{D}$. The census bureau outputs a *sanitized database* (SDB), possibly containing a number of samples different from n. A frequent suggestion is for the census bureau to learn $\mathcal{D}$ and then to publish artificial data obtained by sampling the learned distribution. We reject this solution because we want our sanitized database to reflect (possibly statistically insignificant) "facts on the ground". For example, if a municipality is deciding where to run a bus line, and a certain geographic region of the city has a higher density of elderly residents, it may make sense to run the bus line through this region. Note that such "blips" in populations can occur even when the underlying distribution is uniform.

1.1 Our Approach

- **Privacy.** We use "cryptography-flavored" definitions for (breach of) privacy and effectiveness of a sanitization procedure, but our techniques are purely statistical. Indeed, our work requires *no* computational assumptions. Our definition is based on the intuition that one's privacy is protected to the extent that one blends in with the crowd. Although we do not believe that our ideal definition can be satisfied, it establishes a framework for comparing the privacy offered by different sanitization techniques. We feel this framework is useful beyond our specific suggestions for sanitization, and hope it will influence the nature of discourse on privacy.
- **Utility.** This may be approached in several ways. One approach is a set of "pointwise" results saying that if such-and-such a test is run on the sanitized data, then the resulting statistics are not too far off from the true statistics on the real database. We have two results of this type, involving clustering. A more ambitious project is to find a way of publishing some information that permits testing of a statistical test T on the sanitized data: if T passes the test then the result of running T on the sanitized data is "close" to the result of running T on the real data.
- **Highly Abstracted Version of the Problem.** We consider a very abstract version of the problem, in which each database point is an unlabeled point in high-dimensional space $\mathbb{R}^d$ (in database terminology, there is no object identifier). Thus we completely ignore meanings associated with attributes. Then, very roughly, two points in the space are similar if they are close (the Euclidean distance between them is small). This is unreasonable: distance between points that differ in two cars and one child is the same as between points that differ in one car and two children. Still, our view is that if we cannot understand this abstraction, then we cannot understand real life, while if privacy can be guaranteed under this abstraction, then we may obtain some insight into handling real-life data.

2 What do *We* Mean by "Privacy"?

Our first step was to search the legal and philosophical literature to find a good English language definition of privacy. Gavison [27] describes privacy in terms of controlling access – to our physical person, and to our information. One phrase, "protection from being brought to the attention of others" struck us as particularly relevant to the census problem. Gavison points out that not only is such protection inherent in our understanding of privacy, but moreover when this is compromised, that is, when we have been brought to the attention of others, it invites further violation of privacy, as now our every move is examined and analyzed. This compelling concept – protection from being brought to the attention of others – matches the common intuition that our privacy is protected to the extent that we blend in with the crowd; moreover, it can be converted into a precise mathematical statement.

2.1 A Formal Definition of Privacy

Recall that under our abstraction of the database privacy problem, the real database (RDB) consists of n unlabeled points in high dimensional space $\mathbb{R}^d$, each drawn independently from an underlying distribution $\mathcal{D}$. Intuitively, one is one's collection of attributes. The census bureau publishes a sanitized Database (SDB), containing some n' points, possibly in a different space.

To specify security of a cryptographic primitive we must specify the power of the adversary and what it means to break the system. Since the goal of our adversary is, intuitively, to "single out" a record in the database, we call the adversary an *isolator*. The isolator takes two inputs – the sanitized database and auxiliary information (the auxiliary information is analogous to the history variable in the definition of zero knowledge). The isolator outputs a single point $q \in \mathbb{R}^d$. This completes the description of the functionality of the adversary. Note that the definition admits adversaries of unbounded computational power. Our results require no complexity-theoretic assumptions[1].

We next define the conditions under which the adversary is considered to have succeeded in isolating. The definition is parameterized with two values: a *privacy threshold* t, intuitively, the size of the "crowd" with whom one wants to blend in, and an *isolation parameter* c, whose use will be clear in a moment. Roughly, c helps to formalize "blending in".

For a given isolating adversary $\mathcal{I}$, sanitized database SDB, and auxiliary input z, let $q = \mathcal{I}(\mathrm{SDB}, z)$. Let δ be the distance from q to the nearest real database point, and let x be an RDB point at distance δ from q. Let $B(p, r)$ denote a ball of radius r around point p. If the d-dimensional ball of radius $c \cdot \delta$ and centered at q contains at least t real database points, that is, if $|RDB \cap B(q, c\delta)| \geq t$, then the adversary *fails* to isolate x. Otherwise, the adversary succeeds.

We will give a slightly more general definition shortly. First we give some intuition for the definition. The adversary's goal is to single someone (*i.e.*, some RDB point) out from the crowd. The most likely victim is the RDB point closest to q; call it x. So q "looks something like" x. On the other hand, if $B(q, c\delta)$ contains at least t RDB points, then q also looks almost as similar to lots of (*i.e.*, $t-1$) other RDB points, so x hasn't really been singled out.

Note that the definition of isolation is a relative one; the distance requirement for success varies according to the local density of real database points. This makes sense: if I name an intersection in New York City, there are perhaps a few hundred people living at or very near the intersection, so my point is "close to" many points (people) in the RDB. In contrast, if I name an intersection in Palo Alto, there are perhaps 10 people living near the intersection[2]. More generally, we have the following definition:

Definition 1. **(c-isolation)** *Let y be any RDB point, and let $\delta_y = ||q - y||_2$. We say that q c-isolates y iff $B(q, c\delta_y)$ contains fewer than t points in the RDB, that is, $|B(q, c\delta_y) \cap \mathrm{RDB}| < t$.*

For any point p (not necessarily in the RDB), we let τ_{p} be the minimum radius so that $B(p, \tau_{\mathrm{p}})$ contains t RDB points. We call this the *t-radius of p*.

Claim 1. *Let $y \in RDB$ be arbitrary, and let q be the point output by the isolator. Let $\delta_y = ||q - y||_2$. If $\delta_y > \tau_{\mathrm{y}}/(c-1)$, then y is not isolated by q.*

Proof. Let $p \neq y$ be one of y's $t-1$ nearest neighbors in the RDB, so $||y - p|| \leq \tau_{\mathrm{y}} < (c-1)\delta_y$. Since, by the triangle inequality, $||q - p|| \leq ||q - y|| + ||y - p||$, we have $||q - p|| < \delta_y + (c-1)\delta_y = c\delta_y$. □

We therefore say y is *safe* whenever the conditions of the lemma are satisfied ($\delta_y > \tau_{\mathrm{y}}/(c-1)$).

Thus, we define privacy in the breach; that is, privacy is compromised when the adversary succeeds in isolating even one real database point; privacy is maintained when isolation does not occur. It is an easy consequence of the definitions that if $q = \mathcal{I}(\mathrm{SDB}, z)$ fails to isolate the nearest RDB point to q, then it fails to isolate even one RDB point.

[1] We do not object to using complexity-theoretic assumptions. We simply have not yet had a need to employ them.

[2] This analogy was suggested to us by Helen Nissenbaum.

3 Requirements for the Sanitizer

Suppose one of us publishes all our information on the web – that is, we publish our RDB point x where the adversary can find it, so that the point is part of the adversary's auxiliary information. Clearly, the adversary can isolate x by setting $q = x$ (in which case $\delta_x = 0$ and $B(q, c\delta_x)$ contains only x – we assume no two points are identical). It seems unfair to blame the sanitizing procedure for this isolation; indeed, there is an adversary simulator that, without access to the sanitized database, can also isolate x, since x is part of the auxiliary information. We are therefore concerned with how much seeing the sanitized database helps the adversary to succeed at isolating even one RDB point. Intuitively, we do not want that seeing the SDB should help "too much". Our notion of "too much" is fairly relaxed. Letting ε denote the probability that isolation may occur, we tend to think in terms of $\varepsilon = 1/100$ or $\varepsilon = 1/1000$. This says that about one out of every 100 or 1,000 sanitized databases created may be vulnerable to an isolation event. Actually, ε is a parameter, and can be a function of d and n. Note that ε cannot depend only on n – otherwise privacy could be improved simply by the introduction of additional points.

More formally, a *database sanitizer*, or simply *sanitizer* for short, is a randomized algorithm that takes as input a real database of some number n of points in $\mathbb{R}^d$, and outputs a sanitized database of some number n' of points, in a possibly different space $\mathbb{R}^{d'}$.

A sanitizer is *perfect* if for all isolating adversaries I, for every distribution $\mathcal{D}$ on real database points, there exists an adversary simulator I' such that for all auxiliary information strings z, the probability that $\mathcal{I}(SDB, z)$ succeeds minus the probability that $I'(z)$ succeeds is small. The probabilities are over the coin tosses of the sanitization and isolation algorithms. More precisely, let ε be a parameter (for example, $\varepsilon = 2^{-d/2}$). We require that for all I there exists an I' such that, if we first pick a real database RDB $\in_R \mathcal{D}^n$, then with overwhelming probability over RDB, for all z and for all $x \in$ RDB,

$$|\Pr[\mathcal{I}(SDB, z) \text{ isolates } x] - \Pr[\mathcal{I}'(z) \text{ isolates } x]| < \varepsilon/n,$$

where the probabilities are over the choices made by $\mathcal{I}$, $\mathcal{I}'$, and the sanitization algorithm.

Remark 1: The final quantification (over $x \in$ RDB) is for the following reason. Suppose the adversary knows, as part of its auxiliary information, some point $y \in$ RDB. For every $x \neq y \in$ RDB, we want x's chances of being isolated to remain more or less unchanged when the adversary is given access to the sanitized database. Thus, if we were to write $|\Pr[\exists x\, \mathcal{I}(SDB, z) \text{ isolates } x] - \Pr[\exists x\, \mathcal{I}'(z) \text{ isolates } x]| < \varepsilon$, then we might have $z = y \in$ RDB, $I'(z) = y$, and $I(\text{SDB}, z)$ could (somehow) isolate a different point $x \neq y \in$ RDB with probability one. This is clearly unsatisfactory.

This is excessively ambitious, and we do not think that a nontrivial perfect sanitizer exists. However, we believe that by specifying the ideal we can begin to articulate the value of specific sanitization techniques. For a given technique, we can ask what can be proved about the types of distributions and auxiliary information for which it can ensure privacy, and we can compare different techniques according to these properties.

An occasional reaction to our formulation of privacy is, "What if the crowd that I blend in with shares some particular sensitive attribute, such as having AIDS, being members of the communist party in NY, etc.? Then any point that looks like mine shares this attribute, so my privacy is inadequately protected." This objection assumes that one is more or less completely characterized by certain types of specific sensitive attributes. We think this is an unreasonable form of stereotyping. We believe that individuals cannot be reduced to a very small number of attributes. The data may not be very high dimensional, but neither is it very low dimensional[3]. Furthermore, we operate under the assumption that the crowd-based notion of anonymity described above is adequate for privacy. There may be situations where such a notion does not adequately describe privacy; we do not deal with those situations.

4 Two Classes of Sanitizers

4.1 Perturbation Sanitizers

Recall that, for $x \in \mathbb{R}^d$, τ_x is the minimum radius such that $B(x, \tau_x)$ contains t RDB points. (x need not be an RDB point.) We begin with three sanitizers that are very similar when the dimension d is large. In

[3]Although we obtain results for truly high dimensional data, we believe that our techniques give guarantees on privacy that are related to the inherent dimensionality of the data.

these sanitizers, $d' = d$ and $n' = n$ (that is, the sanitized database consists of points in the same space as the real database, and the numbers of points in the real and sanitized databases are identical). As before, let $B(p, r)$ denote the ball of radius r around p, let $S(p, r)$ denote the corresponding sphere, or the surface of $B(p, r)$. Let $\mathcal{N}(\mu, \sigma^2)$ denote a d-dimensional Gaussian with mean μ and variance σ^2 in every dimension.

1. **The Ball Sanitizer:** For $x \in RDB$, $\mathrm{BSan}(x, RDB) \in_R B(x, \tau_{\mathrm{x}})$.
2. **The Sphere Sanitizer:** For $x \in RDB$, $\mathrm{SSan}(x, RDB) \in_R S(x, \tau_{\mathrm{x}})$.
3. **The Gaussian Sanitizer:** For $x \in RDB$, $\mathrm{GSan}(x, RDB) \in_R \mathcal{N}(x, {\tau_{\mathrm{x}}}^2/d)$.

Since these distributions are very similar, in proving statements about the sanitized database we use whichever sanitization makes the proof easiest. We sometimes refer to these as *round* sanitizers, because of the shape of the distribution on the sanitization of x, given x. These work well for our utility results and can give privacy in certain settings.

In [30], Kargupta et. al. argue that independent additive perturbation may have limited application to preserving privacy insofar as certain informative features of the data set (e.g.: the principal components) are largely unaffected by such perturbations. Their argument assumes (critically) that the sanitizer applies a fixed distribution to each element, and ultimately describes how to reconstruct the covariance matrix of the *attributes*. In this work, we apply data-driven distributions to the elements, and prove privacy guarantees for the *users*. Moreover, we exploit what Kargupta et. al. perceive as a weakness (reconstructing the covariance matrix) to enable a reasonable spectral clustering algorithm, while provably maintaining privacy.

Remark 2: A natural question is how to cope with outliers. One possibility is to suppress them. In some sense, our sanitization procedure does this automatically, as an outlier will have a large t-radius, and so will get sanitized to the point of being meaningless. Note that the existence of an outlier may be disclosed. This does not conflict with our notion of privacy.

4.2 Histogram Sanitizations

Consider some partition of our space $\mathbb{R}^d$ into disjoint *cells*. The histogram for some dataset is a list describing how many points from the dataset fall into each cell. A sanitization procedure is *histogram-based* if it first computes the histogram for the dataset and bases the output only on that information. Matching our intuition, these have strong privacy properties. We have not yet investigated utility under histogram sanitizations. Rather, we are exploring their use in conjunction with round sanitizations in a cross-training setting (see Section 5.1.4 below).

5 Summary of Results

5.1 Privacy

5.1.1 2 Points

We begin with a simple case: the underlying distribution $\mathcal{D}$ is the unit sphere in d dimensions. The database consists of only two points, x and y, which are sanitized with respect to each other. That is, $t = 2$ and each of x and y is perturbed using SSan with perturbation radius $||x - y||$. The adversary is given $||x - y||$, and the sanitizations x' and y'. We show that the probability of 4-isolation is exponentially small in d for all but an exponentially small fraction of choices of the pair (x, y). The proof is a symmetry argument. We show the existence of an exponentially large number of "decoy" pairs (x_i, y_i) for which the posterior probabilities[4] are identical to the posterior probability of (x, y). In addition, we show these are "spread out", so that an adversary can isolate at most one point or decoy point.

[4]The posterior probability at a point p is the probability, given the adversary's view of the world – the priors and the SDB, that x is at point p.

5.1.2 $n - 1$ Points in the Clear and One Sanitized Point

We next consider the case in which the real database consists of n sanitized points, drawn from the d-dimensional unit sphere. The adversary is given all but one point in the clear, together with a sanitization of the final point using SSan. The adversary's goal is to 4-isolate the last point. This is our most interesting proof, and it is sketched in Section 7.1; details appear in the Appendix. As in the two-point case, privacy holds, intuitively, because the hidden point can lie in any direction from its sanitization, while any point q produced by the adversary can only isolate points lying in an exponentially small fraction of these directions.

Remark 3: Unfortunately, it is *not* the case that seeing the sanitizations of the remaining $n - 1$ points, rather than their exact values, gives less information. The reason is that the sanitization of, say, y, implicitly contains information about the t-neighborhood of y. So if x is in the t-neighborhood of y, for example, if x is the $(t - 1)$th nearest neighbor of y, so that $||x - y|| = \tau_{\mathrm{y}}$, then revealing $y' = \mathrm{SSan}(y)$ will leak information about x. We have not yet proved or disproved the viability of the above-mentioned sanitizations.

Before describing our approach to coping with this difficulty, we must discuss our results on histogram sanitizations.

5.1.3 Privacy Results for Histogram Sanitizations

Our simplest histogram result is for the case in which the real database points are distributed uniformly in the cube of side length 2. The cube is subdivided into 2^d sub-cubes of volume 1 in the natural way. The sanitization is the *exact* number of points in each sub-cube. Under these conditions, we show that the probability of 5-isolation is exponentially small in d.

We next consider a refinement of this basic histogram construction, in which whenever a sub-cube contains at least $2t$ points, it is further subdivided into 2^d cubes. We show, as before, that the probability of 8-isolation is exponentially small in d. Note that although we obtain a slightly worse constant $c = 8$ in this case, regardless of the number of times cubes are subdivided, the privacy guarantee stays the same – 8-isolation is exponentially unlikely.

We have several ideas for future work on histogram sanitizations; the one we find most intriguing is described in Section 8.

5.1.4 Putting the Parts Together: Cross Training

A combination of histogram-based sanitization and spherical sanitization gives us a general result on privacy for the case of $n = 2^{o(d)}$.

Our generic approach is as follows. We first randomly partition the real database into two sets A and B. We construct a histogram for B and sanitize A with respect to B; that is, for each $v \in A$, we compute the perturbation radius for elements v by considering only the points in $\{v\} \cup B$. We release the sanitizations of A as well as the exact count of points in A and B in each cell in the histogram for B. In order to protect the privacy of points in B, we ensure that the perturbation radii for points in A are functions of histogram information for B. Since histogram sanitizations are safe, the points in B are safe.

In particular, we consider the following approach: For $v \in A$, let C be the cell in the recursive decomposition of B that contains v. Let ρ_v be the side-length of C. We sanitize v using GSan with the length ρ_v as the standard deviation in every dimension. We show that when $n = 2^{o(d)}$ and the underlying distribution is uniform over a cube, given this sanitization, the probability of 18-isolation is exponentially small in d. We believe that this sanitization, or a slight variant of it, is safe for $n = 2^{\omega(d)}$ as well, and are currently working on proving it.

5.1.5 Very General Results on Round Sanitizers

Finally, we obtain some general results on round sanitizers, for arbitrary distributions. In particular, we show that sanitization can't be made arbitrarily safe: for any distribution, if sanitization is done using BSan, then there is a polynomial time adversary $\mathcal{I}$ requiring no auxiliary information, such that the probability that the adversary succeeds is $\Omega(exp(-d)/\log(n/t))$. Note that our results on privacy described above limit the probability of success for the adversary to $2^{-\Omega(d)}$, nearly matching this lower bound.

We also show that, if the adversary is constrained (in a somewhat unrealistic way), then sanitization can't be too bad: for any distribution, if sanitization is done using BSan, if the adversary is restricted to choosing

$q \in$ SDB, and if the number of points is sufficiently small, and the adversary has no auxiliary information, then the probability of isolation is exponentially small in d.

5.2 Results on Utility

We have two utility results for round sanitization.

1. When the underlying distribution $\mathcal{D}$ is a mixture of Gaussians: There is an efficient algorithm that learns $\mathcal{D}$ from the Gaussian sanitization. Learning mixtures of Gaussians has already been heavily investigated [3, 10, 36]. In our case, the problem is complicated by the fact that noise is added to each sample, and the amount of noise differs from sample to sample, as the amount of noise is determined by the t-radius of the sample.

2. For any distribution, given an algorithm to find k clusters minimizing the maximum diameter of a cluster, if we run the algorithm on data sanitized with either BSan or SSan, then the maximum diameter of a cluster exceeds the maximum diameter of an optimal k-clustering on the RDB by at most a factor of 3.

6 Recent Related Work

Dinur and Nissim [14] investigated *output perturbation*: a query interceptor receives queries, finds the true answer, and adds noise to the true answer before responding (with the noisy answer). They investigated single-attribute binary databases and showed that the amount of noise that had to be added to preserve privacy rendered the database almost useless. Together with Dwork, they also considered the case in which the number of queries is sub-linear in the number of records in the database – a realistic restriction in the case of very large databases – and found that significantly less noise was needed in order to protect privacy. Note that the output perturbation paradigm differs from the sanitization paradigm: in output perturbation, the query interceptor adds fresh noise to the response to each query (this is why it is dangerous to permit too many queries – the noise can cancel); in sanitization, a noisy version of the database is constructed and published once and for all. Evfimievsky, Gehrke, and Srikant [20] investigated sanitization of a transaction database. Dwork and Nissim [19] continued the investigation of *SuLQ* (sublinear query) databases.

The definitions of privacy in [14, 19, 20] all follow more or less the same paradigm: for every record in the database, the adversary's confidence in the values of the given record does not significantly increase as a result of interacting with the database. For example, [14] deals with databases in which each record i consists of a single bit $d_i = b$. The assumption is that the adversary can name individual records in the database; this captures the setting in which the database is really multi-attribute, and the adversary has somehow, out of band, figured out enough about some individual to construct a query that effectively names the individual. Even in such a setting, it should be impossible to learn the value of even a single additional binary data field.

In contrast, our definition is concerned with preventing an adversary from learning strongly identifying information to begin with: intuitively, for any δ, if an isolating adversary can come within distance δ any real database point x, then either this point "looks like" t other points in the database (that is, there are at least $t-1$ other points within distance $c\delta$) or there is an isolating adversary simulator that could isolate x without access to the sanitized database.

7 Two Proof Sketches

The proofs of the results discussed in Section 5 are in the Appendix. We sketch here two of these: privacy for the case of $n-1$ points in the clear and a single sanitized point, and a clustering algorithm that operates on sanitized data.

7.1 Sketch of a Privacy Proof

In this section, we prove that the sanitization SSan preserves privacy of a point, even when all the other points in the dataset are revealed to the adversary. In particular, we perturb only one point in the database, and reveal its sanitization along with the rest of the points to the adversary.

We assume that the dataset consists of n points drawn uniformly and independently from the surface of a sphere S of radius ρ, that is, RDB $= \{x_1, \cdots, x_n : x_i \in_R S\}$. We sanitize the point x_1 with respect to the rest of the points $\{x_2, \cdots, x_n\}$ using a slightly modified version of SSan and with $t = 2$. Recall that the Sphere Sanitizer picks a point $x' \in_R S(x_1, \tau_{x_1})$, where τ_{x_1} is the 1-radius of x_1 in $\{x_2, \cdots, x_n\}$. In this section we assume that x' is picked uniformly from $S \cap S(x_1, \tau_{x_1})$, that is, the new point x' also lies on the surface of the sphere S and is at distance exactly τ_{x_1} from x_1.

The adversary is given S, x' and $\{x_2, \cdots, x_n\}$. Note that since the adversary already knows all the real database points except x_1, her goal is to 4-isolate x_1. Let $D = \{x_2, \cdots, x_n\}$. We study here the case $t = 2$. While we suspect that privacy holds even for larger values of t, this is reserved for future work.

7.1.1 A flat prior

We begin by proving privacy in a somewhat easier setting, and then extend it to the above setting. Assume that points are drawn uniformly at random from space, that is, all points in space are equally likely to be picked in the dataset. This is an unrealistic setting, however, it allows us to reason about privacy without bothering about the underlying distribution, and so is a natural first step.

Given x' and D, let Z be the set of the "legal" points z where x can lie. These are all points z having a 1-radius with respect to D equal to $||z - x'||$.

$$Z = \{z : ||x' - z|| = \min_{y \in D} ||y - z||\}$$

Note that Z is the surface of the convex polyhedron given by $P = \{z : ||x' - z|| \leq ||y - z|| \ \forall y \in D\}$. Let q be any arbitrary point. Define $Q = \{z : ||z - q|| \leq \frac{1}{3}||z - x'||\}$. That is, Q is the set of points 4-isolated by q.

For any set Y, let $F(Y) = \mathbf{Pr}[x \in Y | D, x']$ be the probability, given D and x', that x lies in the set Y. Note that $F(Q)$ is the probability that the point q 4-isolates x_1. Our main aim is to show that regardless of q, $F(Q)$ is exponentially small in d, and therefore, the probability of success for the adversary is exponentially small in d.

Note that $F(Z) = 1$ by definition. Let $G(Y) = \mathbf{Pr}[x', x \in Y | D]$ be the probability density of picking x in the set Y and then perturbing it to x'. By Bayes rule, we have that $F(Y) = G(Y)/\mathbf{Pr}[x'|D]$. Therefore, using $F(Z) = 1$, we get that $F(Y) = \frac{G(Y)}{G(Z)}$.

$G(Y)$ is given by $\int_{z \in Y \cap Z} \frac{1}{A(||z - x'||, d)} f(z) dz$, where $f(z)$ is the probability density of the event that x was picked to be at point z, and $A(\delta, d)$ is the surface area of a sphere of radius δ in d dimensions. Note that $f(z)$ is independent of z, because x is picked from a flat prior. Furthermore, $A(||z - x'||, d) = \Omega_d ||z - x'||^{d-1}$, where Ω_d is the full solid angle in d-dimensions[5]. Thus, we have $G(Y) = \frac{k}{\Omega_d} \int_{z \in Y \cap Z} \frac{1}{||z - x'||^{d-1}} dz$, for some constant k.

Observe that as a set Y gets farther from x', its measure decreases as $\frac{1}{r^{d-1}}$, where r is the distance of the set from x'. However, on the other hand, if the angle subtended by this set at x' remains the same, then the volume of the set increases at the same rate r^{d-1}. In fact we show that the measure $G(Y)$ is proportional to the solid angle subtended by the set Y at the point x' (see proof of Lemma 2). The following two lemmas show that while the set Z subtends a large angle at x', Q subtends an exponentially small angle. Note that if the polygon P is bounded, the set Z subtends the full solid angle at x'. The first lemma follows by showing that in the worst case, the set Z spans at least half of the full solid angle at x'. Proofs are given in appendix A.2.

The intuition behind the lemmas is that the set Z has a high "directional-spread". That is, the pre-image of x' could lie in many possible directions. However, the adversary can only isolate points in one general direction with respect to x. Note that this is very similar to the existence of many different "decoys" for the 2-point case described in Section 5.1.1.

Lemma 2. $G(Z) \geq \frac{1}{2}k$.

Lemma 3. $G(Q) \leq 2^{1-d}k$.

We get the following theorem:

Theorem 4. $F(Q) \leq 2^{2-d}$.

[5]The full solid angle in d-dimensions is equal in measure to the surface area of a d-dimensional sphere of radius 1.

7.1.2 Points drawn from the surface of a sphere

Now we extend the above lemmas to the setting where points are drawn uniformly from the surface of a sphere S of radius ρ. In order to get results analogous to the previous subsection, we assume that the ambient space is $(d+1)$-dimensional, and the surface of the sphere is d-dimensional. Consider sets Z, P and Q defined as before, but on the surface of the sphere S. Measures F and G are defined as before. Recall that our goal is to bound $F(Q) = \frac{G(Q)}{G(Z)}$. Note that because the probability density of the underlying distribution is uniform over the surface, as before, for any set Y on the surface of the sphere, we have $G(Y) = \frac{k}{\Omega_d}\int_{z\in Y\cap Z}\frac{1}{||z-x'||^{d-1}}dz$, for some constant k. This holds in particular for the sets Z and Q.

On the surface of the sphere, the polygon P is always bounded. However, because it lies on a curved surface, this does not imply that the set Z subtends the full d-dimensional angle at x'. Instead we get the following weaker statement:

Lemma 5. (Counterpart to Lemma 2) $G(Z) \geq 2^{-(1+d/2)}k$.

The measure of $Q \cap P$ on the other hand can be bounded in a similar way as in Lemma 3. This implies the following lemma.

Lemma 6. (Counterpart to Lemma 3) $G(Q) \leq 2^{1-3d/4}k$.

Lemmas 5 and 6 give the following theorem:

Theorem 7. $F(Q) \leq 2^{2-d/4}$. *Therefore, the probability that an adversary succeeds in isolating x_1 based on her knowledge of x', $x_2, \cdots, x_n$ and the sphere S, is at most $2^{2-d/4} = 2^{-\Omega(d)}$.*

7.2 Sketch of a Utility Proof

In this section we look at an algorithm for mining sanitized data. We work with the fairly common problem of learning a mixture of Gaussians. We assume there are k mean vectors μ_i, each with an associated mixing weight w_i. Let $w_{\min}$ denote the minimum of the mixing weights. Each point in the data set is independently produced by selecting a μ_i with probability proportional to the w_i, and applying independent, normally distributed error to each coordinate. We assume that the Gaussians are spherical, in that Gaussian has an associated variance that is used for each of the coordinates, and that the variance of each is at most σ_1^2. We assume that the sanitization process amounts to applying an additive perturbation established by choosing a point uniformly at random from the unit sphere, which is then scaled in length by a random variable at most $2\sigma_2\sqrt{d}$, that may *depend on the sampled data*. Notice that this is sufficiently general to capture all the perturbation based sanitizations described in Section 4.1 – SSan, BSan, and GSan.the latter two using a random variable for the scaling factor.

For the purposes of analysis, we assume that we have access to two data sets $\overline{A}$ and $\overline{B}$ that have been independently sanitized. Each is assumed to result from the same underlying set of means, and to be independent of the other. We use $\overline{A}_u$ to denote the sanitized vector associated with u, $\widehat{A}_u$ the unsanitized vector associated with u, and A_u the original mean vector associated with u. We form the matrices $\overline{A}$, $\widehat{A}$, and A by collecting these columns for each point in the data set. Let w_u denote the mixing weight associated with the Gaussian with mean A_u. The matrices $\overline{B}$, $\widehat{B}$, and B and their associated columns are analogous, though they represent an independently sanitized disjoint data set. While this setting is not difficult for the sanitization process to accommodate, the motivation is simply for the clarity of analysis and it is unclear that disjoint data sets are necessary in practice.

7.2.1 Preliminaries:

We first remind the reader of the matrix L_2 norm, written and defined as:

$$\|M\|_2 \quad = \quad \max_{x\neq 0}|Mx|/|x| \tag{1}$$

The main linear algebraic tool that we use is a matrix's *optimal rank k projection.* Each projection matrix is associated with a subspace, and performs the linear operation which "flattens" points to the subspace. For every matrix M, there is an optimal rank k projection matrix P_M, such that for all rank k matrices D

$$\|M - P_M M\|_2 \quad \leq \quad \|M - D\|_2 \tag{2}$$

Computing these optimal subspaces is not difficult; in most cases it is an $O(dn\log(dn))$ operation.

We also make use of the *single linkage* clustering algorithm [32]. For our purposes, given a collection of points single linkage repeatedly inserts an edge between the closest pair of non-adjacent points until the resulting graph has k connected components.

7.2.2 Algorithm:

Using optimal rank k projections and single linkage clustering, we now present our algorithm.

Cluster$(\overline{A}, \overline{B}, k)$

1. Compute $P_{\overline{A}}$ and $P_{\overline{B}}$, and form $C = [P_{\overline{B}}\overline{A} | P_{\overline{A}}\overline{B}]$.
2. Apply single linkage to C, forming k clusters.

Cluster takes a pair of data sets, and uses the structure define by each data set to filter the noise from the points in the other. If the mean vectors μ_i have sufficient separation, all inter-cluster distances in C will exceed all intra-cluster distances in C, and single linkage will associate exactly those points drawn from the same mean.

Theorem 8. *Assume that $d < n/2$. If for each pair of means μ_i, μ_j,*

$$|\mu_i - \mu_j| \quad \geq \quad 4(\sigma_1 + \sigma_2)(16/w_{\min}^{1/2} + \sqrt{k\log(kn/\delta)})$$

*then with probability $1 - 2(e^{-nw_{\min}/8} + 2^{-log^6 n/2} + \delta)$, **Cluster** partitions the columns of $\overline{A}$ and $\overline{B}$ according to the underlying Gaussian mixture*

Proof. The proof is conducted by bounding $|\mu_u - C_u|$ for each u. Assume, without loss of generality, that $\mu_u = A_u$ and $C_u = P_{\overline{B}}\overline{A}_u$. Notice that

$$\begin{aligned} |\mu_u - C_u| &= |A_u - P_{\overline{B}}\overline{A}_u| \\ &\leq |A_u - P_{\overline{B}}A_u| + |P_{\overline{B}}A_u - P_{\overline{B}}\overline{A}_u| \end{aligned}$$

Lemmas 30 and 31 (see Appendix D.1) bound the norms of these two terms, respectively, so that their sum is at most 1/4 the assumed separation of the mean vectors. With such separation, all inter-cluster distances are at least 1/2 the mean separation, and all intra-cluster distances are at most 1/2 the mean separation. □

Although the above lemma requires a uniform bound on the pairwise separation between the means of the clusters, by using a more sophisticated clustering algorithm than Single-Linkage on the low-dimensional projection of the data, we can improve the results such that the requisite separation between a pair of means depends only on the variances of the corresponding Gaussians. We omit the details.

8 Summary and Future Work

The immediate focus of future work will be to investigate histogram sanitizations in the context of the "round" (spherical, ball, and Gaussian) distributions. There are several natural ideas; one of our favorites is to first create Voronoi regions induced by the origin together with some number m of new samples from the distribution (these samples are independent of the data) and publish (possibly approximate) histogram information (of the real data) for this partitioning of the space. As above, for any Voronoi cell containing sufficiently many points, we would recurse. The advantage to this approach over cubic histograms is that the histogram regions will better match the data; we believe this will give improved utility over the results one might obtain by cross-training in which the sample space is round but the histogram is based on cubes.

Another pressing direction is to further explore utility, in particular, a method for assessing the validity of results obtained by applying a given statistical test to sanitized data.

An important aspect of any sanitization technique is to consider its application in an online setting, where the database changes over time. We feel that sanitizations of points should not be recomputed independently as the database changes, because an adversary collecting information over time may be able to gather enough to filter out the noise. However, in situations such as when one of the t-nearest neighbors of a point dies, one may be forced to recompute the sanitization. We believe that in such a case the new sanitization should be conditioned on the previous one appropriately, so as to prevent leakage of extra information. A related open area is to extend the definitions and techniques to multiple databases.

Then, there are the more obvious questions: how to cope with discrete data, or even non-numerical data. In general, to draw a connection to real life data, we will need to scale different attributes appropriately, so that the data are well-rounded. This requires some formal treatment.

We have learned several lessons from the abstract model. In particular,

1. High dimensionality is our friend: our results typically show that the probability of isolation is exponentially small in the dimension d.

2. Gaussian perturbations, where the perturbation radius is determined by local density information, seem to be reasonable, both from the privacy and the utility perspectives.

3. Outliers are treated automatically by the sanitization procedure: outliers are highly randomized, essentially resulting in their obliteration.

Acknowledgements. We have benefited enormously from numerous conversations with many people, and are particularly grateful to Dimitris Achlioptas, Jin-Yi Cai, Petros Drineas, John Dunagan, Amos Fiat, Michael Freedman, Russel Impagliazzo, Michael Isard, Anna Karlin, Moni Naor, Helen Nissenbaum, Anna Redz and Luca Trevisan. Finally, we thank Larry Stockmeyer for many contributions, including Theorems 20, 21 and 34.

References

[1] D. Agrawal and C. Aggarwal, On the Design and Quantification of Privacy Preserving Data Mining Algorithms, *Proceedings of the 20th Symposium on Principles of Database Systems*, 2001.

[2] N. R. Adam and J. C. Wortmann, Security-Control Methods for Statistical Databases: A Comparative Study, *ACM Computing Surveys* 21(4): 515-556 (1989).

[3] S. Arora and R. Kannan. Learning mixtures of arbitrary Gaussians. *ACM STOC*, 2001.

[4] R. Agrawal and R. Srikant, Privacy-preserving data mining, *Proc. of the ACM SIGMOD Conference on Management of Data*, pp. 439–450, 2000.

[5] Beck, L., A security machanism for statistical database, *ACM Transactions on Database Systems (TODS)*, 5(3), p.316-3338, 1980.

[6] Cox, L. H., Suppression Methodology and Statistical Disclosure Control, *Journal of the American Statistical Association*, Vol. 75, 1980, pp. 377-385.

[7] Cox, L. H., Solving Confidentiality Protection Problems in Tabulations Using Network Optimization: A Network Model for Cell Suppression in the U.S. Economic Censuses, *Proceedings of the International Seminar on Statistical Confidentiality*, International Statistical Institute, Dublin, 1982, pp.229-245.

[8] Cox, L. H., Disclosure Control Methods for Frequency Count Data, presented at the Census Advisory Committee Meeting of the American Statistical Association, Bureau of the Census, 1984.

[9] Cox, L. H., New Results in Disclosure Avoidance for Tabulations, *International Statistical Institute Proceedings of the 46th Session*, Tokyo, 1987, pp. 83-84.

[10] S. Dasgupta, Learning mixtures of Gaussians, *IEEE FOCS*, 1999.

[11] Denning, D. E., Schlorer, J., and Wehrle, E., Memoryless inference controls for statistical databases. Computer Science Dept., Purdue University, 1982.

[12] Denning, D., Secure statistical databases with random sample queries, *ACM Transactions on Database Systems (TODS)*, 5(3), p.291-315, 1980.

[13] P. Diaconis and B. Sturmfels, Algebraic Algorithms for Sampling from Conditional Distributions, *Annals of Statistics 26*(1), pp. 363–397, 1998

[14] I. Dinur and K. Nissim, Revealing information while preserving privacy, *Proceedings of the Symposium on Principles of Database Systems*, pp. 202-210, 2003.

[15] A. Dobra and S.E. Fienberg, Bounds for cell entries in contingency tables given marginal totals and decomposable graphs, *Proceedings of the National Academy of Sciences 97*, No. 22, pp. 11885–11892, 2000

[16] A. Dobra and S.E. Fienberg, Bounding entires in multi-way contingency tables given a set of marginal totals, *Foundations of Statistical Inference: Proceedings of the Shoresh Conference 2000*, pp. 3–16, Springer-Verlag, 2003.

[17] A. Dobra and S.E. Fienberg, and M. Trottini, Assessing the risk of disclosure of confidential categorical data, *Bayesian Statistics 7*, pp. 125–14, Oxford University Press, 2000.

[18] G. Duncan, Confidentiality and statistical disclosure limitation. In N. Smelser & P. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences*. New York: Elsevier. 2001

[19] C. Dwork and K. Nissim, Privacy-Preserving Datamining on Vertically Partitioned Databases, *Manuscript (submitted for publication),* 2004.

[20] A. V. Evfimievski, J. Gehrke and R. Srikant, Limiting privacy breaches in privacy preserving data mining, *Proceedings of the Symposium on Principles of Database Systems*, pp. 211-222, 2003.

[21] Fellegi, I. P., On the question of statistical confidentiality. *J. Am. Stat. Assoc.* 67:337 (Mar.), 1972, pages 7-18.

[22] W. Feller. An Introduction to Probability Theory & Its Applications, Volume 1. Wiley Text Books, 3rd edition, 1968.

[23] S. Fienberg, Confidentiality and Data Protection Through Disclosure Limitation: Evolving Principles and Technical Advances, IAOS Conference on Statistics, Development and Human Rights September, 2000, available at

`http://www.statistik.admin.ch/about/international/fienberg_final_paper.doc`

[24] S. Fienberg, U. Makov, and R. Steele, Disclosure Limitation and Related Methods for Categorical Data, *Journal of Official Statistics, 14*, pp. 485–502, 1998.

[25] L. Franconi and G. Merola, Implementing Statistical Disclosure Control for Aggregated Data Released Via Remote Access, Working Paper No. 30, *United Nations Statistical Commission and European Commission, joint ECE/EUROSTAT work session on statistical data confidentiality*, April, 2003, available at `http://www.unece.org/stats/documents/2003/04/confidentiality/wp.30.e.pdf`

[26] Füredi, Zoltán and Komlós, János, The eigenvalues of random symmetric matrices, *Combinatorica*, 1:3, 1981, pages 233–241.

[27] R. Gavison, Privacy and the Limits of the Law, in Deborah G. Johnson and Helen Nissenbaum, editors, *Computers, Ethics, and Social Values*, pp. 332–351. Prentice Hall, 1995.

[28] D. Gusfield, A graph Theoretic Approach to Statistical Data Security, *SIAM Journal on Computing 17*(3), pp. 552–571, 1988

[29] P. Indyk and R. Motwani. Approximate Nearest Neighbor: Towards Removing the Curse of Dimensionality. *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, 1998.

[30] Kargupta, Hillol and Datta, Souptik and Wang, Qi and Sivakumar, Krishnamoorthy, On the Privacy Preserving Properties of Random Data Perturbation Techniques, *Proceedings of the Third ICDM IEEE International Conference on Data Mining*, 2003.

[31] J. M. Kleinberg, C. H. Papadimitriou, and P. Raghavan, Auditing Boolean Attributes, *J. Comput. Syst. Sci. 66*(1), pp. 244–253, 2003

[32] Sibson, R, SLINK: an optimally efficient algorithm for the single-link cluster method, In *the Computer Journal* Vol. 16, No. 1, 1973, pages 30–34.

[33] A. Shoshani, Statistical databases: Characteristics, problems and some solutions, *Proceedings of the 8th International Conference on Very Large Data Bases (VLDB'82)*, pages 208–222, 1982.

[34] A. Slavkovic and S.E. Fienberg, Bounds for Cell Entries in Two-way Tables Given Conditional Relative Frequencies, Submitted to *Privacy in Statistical Databases '2004*, Barcelona, June 2004.

[35] L. Sweeney, k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10 (5), 2002; 557-570.

[36] S. Vempala and G. Wang, A spectral algorithm for learning mixtures of distributions, *IEEE FOCS*, 2002.

A Perturbation Sanitizations

A.1 Two Points

Theorem 9. *Suppose the underlying distribution is uniform on a sphere of radius ρ in $\mathbb{R}^d$, and the real database consists of two points x, y with $||x-y|| = \delta$. In addition, the isolating adversary has auxiliary information δ as well as the description of the sphere (the center and radius ρ). Then, if sanitization is done using* SSan *with $t = 2$ and if $\delta < \left(2\sqrt{\frac{6}{7}} - \epsilon\right)\rho$ (for any constant $\epsilon > 0$)*[6]*, every isolating adversary succeeds at 4-isolating a point with probability at most $2^{-\Omega(d)}$.*

Note that if we pick two points x, y uniformly at random from the underlying spherical distribution[7],

$$\Pr\left[\delta < \left(2\sqrt{\frac{6}{7}} - \epsilon\right)\rho\right] > 1 - 2^{-\Omega(d)}$$

Therefore, security is guaranteed for an overwhelming fraction of inputs from the distribution.

Proof. Let x', y' denote the sanitizations of points x and y respectively. We begin by specifying the key lemma underlying this proof:

Lemma 10. *For fixed x and y, with probability at least $1-p$ over the choice of x' and y', there exist $N = 2^{\Omega(d)}$ distinct "decoys" x_i, with $x_i \neq x$, that have the following properties:*

- *(identical priors) Given x', y', δ, the real data point x is equally likely to be any of the points $x, x_i, \ldots, x_N$. That is, for all i, $\Pr[x \in B(x_i, \frac{\delta}{3}) \mid x', y', \delta] = \Pr[x \in B(x, \frac{\delta}{3}) \mid x', y', \delta]$.*
- *(well-spread) For all $i \neq j$, $B(x_i, \frac{\delta}{3}) \cap B(x_j, \frac{\delta}{3}) = \emptyset$. For all i, $B(x_i, \frac{\delta}{3}) \cap B(x, \frac{\delta}{3}) = \emptyset$.*

This is how we use "decoys" in our proof of privacy: given the adversary's view of the database (namely x', y', o, δ), all of the decoys are indistinguishable and equally likely as the original inputs as x itself. Furthermore, the well-spread property guarantees that any point that the adversary may output would 4-isolate only at most one of the decoys. It follows that the adversary cannot 4-isolate x with probability greater than $2^{-\Omega(d)}$. Applying the union bound over x and y, we have:

$$\begin{aligned} & \Pr[\text{the adversary 4-isolates } x \text{ or } y] \\ \leq\ & \Pr[\text{the adversary 4-isolates } x] + \Pr[\text{the adversary 4-isolates } y] \\ \leq\ & 2^{-\Omega(d)} \end{aligned}$$

□

[6] or equivalently, $\angle xoy < \pi - 2\sin^{-1}\sqrt{2/7} - \epsilon$

[7] The more general statement is as follows: for all $\epsilon > 0$, $\Pr_{x,y}[\,|\cos\angle xoy\,| > \epsilon] < 2^{-\epsilon^2 d}$.

We construct decoys by reflecting x and y through many $(d-1)$-dimensional hyperplanes through x', y' and o. In order to show that the reflections are sufficiently far from x and y, we need the following technical result stating that with high probability over the sanitization, x is sufficiently far from the 2-dimensional hyperplane $H_2(o, x', y')$ passing through o, x' and y'.

Claim 11. *With probability at least $1 - p = 1 - 2^{-\Omega(d)}$ over the choice of x' and y',*

$$|x - H_2(o, x', y')| > \left(\frac{\sqrt{2}}{3} + \epsilon'\right)\delta \tag{3}$$

(where $\epsilon' > 0$ is a constant that depends on ϵ).

In the rest of our analysis, we will focus on the case where $|x - H_2(o, x', y')| > \left(\frac{\sqrt{2}}{3} + \epsilon'\right)\delta$, which happens with probability at least p.

Claim 12. *Suppose $|x - H_2(o, x', y')| > \left(\frac{\sqrt{2}}{3} + \epsilon'\right)\delta$. Then, there exists $N = 2^{\Omega(d)}$ $(d-1)$-dimensional hyperplanes H_i through x', y' and o, such that for all i, $|x - H_i| > \frac{\delta}{3}$, and for all $i \neq j$, the angle between H_i and H_j is at least $\frac{\pi}{3}$.*

Proof of Claim 12: Let $\hat{h}_x$ denote the normal vector from x to the plane $H_2(o, x', y')$, so $|x \cdot \hat{h}_x| > \sqrt{2}\delta/3$. Choose $\hat{h}_1, \ldots, \hat{h}_{d-3}$ to be orthogonal unit vectors that are normal to the plane $H_2(o, x', y')$ and to the vector $\hat{h}_x$. Now consider the following collection of vectors.

$$\Gamma = \left\{ \alpha\hat{h}_x + \beta\hat{h} \;\middle|\; \hat{h} \text{ is a unit vector spanned by } \hat{h}_1, \ldots, \hat{h}_{d-3} \right\}$$

where $\alpha = \frac{1}{\sqrt{2+3\epsilon'}} = \frac{1}{\sqrt{2}} - \Omega(1)$ and $\beta = \sqrt{1-\alpha^2} = \frac{1}{\sqrt{2}} + \Omega(1)$. Then, any vector $\hat{\gamma} \in \Gamma$ is a unit vector that defines a $(d-2)$-dimensional plane $\pi_\gamma = \{u \mid u \cdot \hat{\gamma} = 0\}$ with the following properties:

1. π_γ contains the plane $H_2(o, x', y')$ (the vectors ox', oy' are orthogonal to γ); and
2. The distance of x from π_γ is given by $\mid x \cdot \hat{\gamma} \mid = \mid \alpha x \cdot \hat{h}_x \mid > \frac{\delta}{3}$.

Observe that Γ defines a sphere of radius β that sits in a $(d-2)$-dimensional subspace of $\mathbb{R}^d$. It is easy to show via a probabilistic argument[8] that there exists $N = 2^{\Omega(d)}$ points $\hat{\gamma}_i$, $i = 1, 2, \ldots, N$ on Γ such that the pairwise distance between any two points is more than 1. This implies that the angle (subtended at o) between any two vectors $\hat{\gamma}_i, \hat{\gamma}_j$, $i \neq j$ is at least $2\sin^{-1}\frac{1}{2} = \frac{\pi}{3}$. We may then pick our hyperplanes to be those given by $H_i = \pi_{\gamma_i} = \{u \mid u \cdot \hat{\gamma}_i = 0\}$. The angle between any two of these hyperplanes is equal to that between their corresponding normal vectors, which is at least $\frac{\pi}{3}$, and claim 12 follows. $\square$

We may now complete the construction of decoys specified in Lemma 10 using the "reflection technique". *Proof of Lemma 10:* For each i, we define $x_i = R_{H_i}(x) = x_i = x - 2(x \cdot \hat{\gamma}_i)\hat{\gamma}_i$, where $R_H(u)$ denotes the image of a point u under reflection in a hyperplane H. Observe that reflection preserves distances (in ℓ_2-norm), and since H_i passes through the points x', y' and o, the points x_i also lie on $B(0, \rho)$. In addition, we also have $||x_i - x'|| = ||x - x'||$ and $||x_i - y'|| = ||x - y'||$. The "identical priors" property of the decoys then follows from a symmetry argument.

Next, we need to verify that the decoys satisfy the "well-spread" property. That $B(x, \delta/3)$ and $B(x_i, \delta/3)$ are disjoint follows from $||x - H_i|| > \delta/3$. To see that $B(x_i, \delta/3)$ and $B(x_j, \delta/3)$ are disjoint for $i \neq j$, consider the lines (x, x_i) and (x, x_j). These lines are normal to the hyperplanes H_i and H_j respectively. Thus the angle between the lines is the same as the angle between the hyperplanes, which is at least $\frac{\pi}{3}$, by claim 12. Since $|x - x_i| > \frac{2}{3}\delta$, and $|x - x_j| > \frac{2}{3}\delta$, we have $|x_i - x_j| > \frac{2}{3}\delta$. Thus, $B(x_i, \frac{\delta}{3}) \cap B(x_j, \frac{\delta}{3}) = \emptyset$. $\square$

[8] The probability that two random points on a sphere of radius β has distance at most 1 is $2^{-\Omega(d)}$. This allows us to take a union bound over pairwise distances amongst $2^{\Omega(d)}$ points.

A.1.1 Proof of Claim 11

Proof of Claim 11 (Sketch): Let ℓ_1 denote the line passing through x and parallel to oy. First, we will show that for some $\epsilon' > 0$ (that depends only on ϵ), with probability $1 - 2^{-\Omega(d)}$ over x',

$$|\ell_1 - H_2(o, x', y)| > \left(\frac{\sqrt{2}}{3} + 2\epsilon'\right)\delta \tag{4}$$

Since $|x - H_2(o, x', y)| \geq |\ell_1 - H_2(o, x', y)|$, it would follow that with probability $1 - 2^{-\Omega(d)}$ over x',

$$|x - H_2(o, x', y)| > \left(\frac{\sqrt{2}}{3} + 2\epsilon'\right)\delta \tag{5}$$

We define $B = \{x' \in S(x, \delta) \mid (4) \text{ does not hold }\}$, and we want to bound the ratio of the measure of B to that of $S(x, \delta)$. To do this, we divide $S(x, \delta)$ into "thin slices" of $(n-1)$-dimensional hyperplanes σ normal to ℓ_1 (and thus oy). For each σ, $B \cap \sigma$ is the union of 2 spherical caps on $S_\sigma = S(x, \delta) \cap \sigma$, where S_σ is a $(d-1)$-dimensional sphere. Now, if σ is close to x (that is, $|x - \sigma| < \epsilon_1$ for some constant ϵ_1), we show that the spherical caps subtend angles smaller than $\pi/2$, and thus the ratio of the areas of $B \cap \sigma$ to that of $S(x, \delta) \cap \sigma$ is $2^{-\Omega(d)}$. On the other hand, we know that with probability $1 - 2^{-\Omega(d)}$ over x', x' lies on some hyperplane σ that is close to x. We can then take a union bound over the two cases. To formalizing the notion of dividing $S(x, \delta)$ into "thin slices", we only have to interpret the computation as an integral in a straight-forward manner. Therefore, we will only focus on the analysis for each σ sufficiently close to x.

First, consider the case when σ passes through x, and let v denote the point $oy \cap \sigma$. Observe that $|xv| = \delta \sin(\pi/2 - \angle xoy/2) = \left(\sqrt{2/7} + \Omega(1)\right)\delta$. Choose $\epsilon' > 0$ sufficiently small so that $\frac{|vx|}{\sqrt{\delta^2 + |vx|^2}} > \frac{\sqrt{2}}{3} + 3\epsilon'$. Consider what happens as we vary x' over S_σ. Then, (4) is satisfied whenever $\angle x'vx \in (\theta, \pi - \theta)$, where $\theta = \sin^{-1} \frac{\delta}{|vx|}\left(\frac{\sqrt{2}}{3} + 2\epsilon'\right)$. The latter holds whenever $\theta_1 < \angle vxx' < \theta_2$ for some θ_1, θ_2 satisfying $\theta_1 < \pi/2 < \theta_2$. It remains to bound the ratio of the areas of spherical caps on S_σ subtending angles θ_1 and $\pi - \theta_2$ to that of the area of S_σ. This ratio is $2^{-\Omega(d)}$ because each of the angles is $\theta/2 - \Omega(1)$.

Next, consider the case where $|\sigma - x| < \epsilon_1$ for some small constant $\epsilon_1 > 0$, and note that S_σ is a sphere in $(d-1)$-dimensional space of radius $\sqrt{\sigma^2 - \epsilon_1^2} \approx \sigma$. As long as ϵ_1 is sufficiently small, it is easy to see that the same analysis as that for σ passing through x works and yields similar bounds.

Next, we show that conditioned upon (5) being true, (3) happens with probability $1 - 2^{\Omega(d)}$ over choice of y'. That is, we want to show:

$$\Pr_{y'}\left[|x - H_2(o, x', y')| > (\sqrt{2}/3 + \epsilon')\delta \;\middle|\; |x - H_2(o, x', y)| > (\sqrt{2}/3 + 2\epsilon')\delta\right] > 1 - 2^{\Omega(d)}$$

Now, suppose $|x - H_2(o, x', y)| > (\sqrt{2}/3 + 2\epsilon')\delta$, and consider the line ℓ_2 passing through x and parallel to ox'. As with the previous proof, we divide $S(y, \delta)$ into "thin slices" of $(n-1)$-dimensional hyperplanes σ' normal to ℓ_2 (and thus ox), and again, we only consider the case where σ' passes through y, as the case where σ' is close to y is similar. Let v' denote the point $ox' \cap \sigma'$. We know that the line $v'y$ does not intersect the sphere $S(x, (\sqrt{2}/3 + 2\epsilon')\delta)$, and we want to know how large $\angle yv'y'$ can be before it intersects the sphere $S(x, (\sqrt{2}/3 + \epsilon')\delta)$. The worst case is when σ' passes through x. In this case, it is easy to see that as long as $\angle yv'y' \leq \theta'$, for some angle $\theta' > 0$ that depends only on ϵ', $v'y'$ does not intersect the sphere $S(x, (\sqrt{2}/3 + \epsilon')\delta)$. This shows that as long as y' does not lie in some spherical cap subtending some angle $\pi/2 - \Omega(1)$ which occurs with probability $1 - 2^{-\Omega(d)}$, we have the required property.

□

A.2 Proofs for Section 7.1

Proof of Lemma 2: The solid angle subtended by a surface P at a point p is given by the integral $\omega_p(P) = \int_P \frac{\hat{\mathbf{n}} \cdot d\mathbf{a}}{r^{d-1}}$, where $d\mathbf{a}$ is the differential area[9] of a surface patch, $\hat{\mathbf{n}}$ is a unit vector from p to the surface patch, and r is the distance of the surface patch from p.

[9]The differential area is a vector with length equal to the area measure of an infinitesimal surface patch and a direction perpendicular to the surface. The formula implies that for any given angle subtended at point p, the minimum surface area that subtends that angle is perpendicular to the ray joining the patch to p, or in other words, belongs to a sphere centered at p.

Note that $\hat{\mathbf{n}} \cdot d\mathbf{a} \leq da$. Therefore, the solid angle $\omega_p(P) \leq \int_P \frac{da}{r^{d-1}} = \int_{z \in P} \frac{dz}{|z-p|^{d-1}}$. Using $p = x'$ and noting that Z is a surface, we get that $G(Z) \geq \frac{k}{\Omega_d}\omega_{x'}(Z)$. In order to complete the proof, we show that $\omega_{x'}(Z) \geq \frac{1}{2}\Omega_d$.

Recall that Z is the boundary of intersection of the half spaces $H_y = \{z : |x' - z| \leq |y - z|\}$, $\forall y \in D$. Consider any one halfspace H_y. Let p be the point on the boundary of this half-space, that is closest to x'. Note that the line segment px' is perpendicular to the hyperplane $\{z : |x' - z| = |y - z|\}$ defining H_y. Let ℓ be any ray starting at x' and making an angle less than $\pi/2$ with the segment px'. The ray ℓ cannot lie entirely within the half-space H_y because it makes an acute angle with px'. Therefore, it does not lie entirely within P, which is an intersection of several half-spaces including H_y, and so, it must intersect the boundary of P, i.e. Z. Note that this holds for all rays ℓ that make an acute angle with px'. This implies that the angle subtended by Z at x' is at least as much as the angle subtended at x' by a d-dimensional hemisphere with center x'. The latter being $\frac{1}{2}\Omega_d$, we get the result. □

Proof of Lemma 3: Let $A(Y)$ denote the surface area of set Y and $V(Y)$ denote its volume. Let $|q - x'| = r$. First note that for all $x \in Q$, $|x - x'| \geq \frac{3}{4}r$, by triangle inequality. Thus, $G(Q) = k \int_{z \in Q \cap Z} \frac{1}{A(|z-x'|,d)} dz \leq k \int_{z \in Q \cap Z} \frac{1}{A(\frac{3}{4}r,d)} dz = kV(Q \cap Z)\frac{1}{A(\frac{3}{4}r,d)}$, where $V(Q \cap Z)$ is the total measure contained in $Q \cap Z$.

In order to bound $V(Q \cap Z)$, we note that Z is the surface of the convex polyhedron P. Thus, $V(Q \cap Z) \leq A(Q \cap P)$. Now we use a fact from geometry that states that the surface area of a convex body contained inside another is bounded by the surface area of the outer body ["Theory of Convex Bodies" by Bonnesen and Fenchel]. This implies that $A(Q \cap P) \leq A(Q)$.

In order to bound $A(Q)$, we note that Q is an ellipsoid with the longest diameter being equal to $\frac{3}{4}r$. Thus, it is contained in a sphere of radius $\frac{3}{8}r$. Consequently, its surface area is bounded above by $A(\frac{3}{8}r, d)$. So we get $G(Q) \leq kV(Q \cap Z)\frac{1}{A(\frac{3}{4}r,d)} \leq k\frac{A(\frac{3}{8}r,d)}{A(\frac{3}{4}r,d)} \leq k2^{1-d}$. □

Proof of Lemma 5: The set Z of all legal pre-images of x' is a $d-1$ dimensional surface on S in a $d+1$ dimensional space. Analogous to the proof of Lemma 2, $G(Z)$ is proportional to the d-dimensional solid angle subtended by Z at x'. We begin with some observations about the set Z. As before, Z is the boundary of intersections of several half-spaces. Note that as the intersection gets smaller (as more and more half-spaces are added), the angle subtended by the boundary gets larger. Therefore, we bound the angle from below by considering the angle subtended by the boundary of a single half-space, which is a hyperplane.

The hyperplane under consideration is the set of all points that are equidistant from x' and some real point $x_i \in D$. Since both these points lie on the surface of S, the hyperplane passes through the center of S and its intersection with the surface is a "great circle" (in d-dimensions) of S. At least half of the great circle is at distance at most $\sqrt{2}\rho$ from x'. Therefore, the angle subtended by the great circle is at least half its area divided by $A(\sqrt{2}\rho, d)$. We get $G(Z) \geq k\frac{1}{2}\frac{A(\rho,d)}{A(\sqrt{2}\rho,d)} = k2^{-(1+d/2)}$. □

Proof of Lemma 6: As before we have $G(Q) \leq kV(Q \cap Z)\frac{1}{A(\frac{3}{4}r,d)} \leq kA(Q \cap P)\frac{1}{A(\frac{3}{4}r,d)}$. Now, $Q \cap P$ is a d-dimensional surface on S. Again as before, it is contained inside the intersection of a sphere of radius $\frac{3}{8}r$ with S. The latter intersection is a spherical cap C of S.

Note that since we are on a curved surface, the theorem about the surface area of a convex body contained inside another does not apply. However, we will use the fact that the curvature of the cap is bounded. In particular, let Π be the orthogonal projection of the cap onto a hyperplane perpendicular to the line joining the center of the sphere S to the center of the cap C. Note that the projection Π always shrinks lengths. Using $r \leq \rho$ and the fact that C has a radius of $\frac{3}{8}r$, some simple geometry shows that Π shrinks lengths by a factor of at most $2^{1/4}$. So for any d-dimensional surface Y inside C, $A(Y) = 2^{d/4}A(\Pi(Y))$.

Consider the projection $\Pi(C)$. This is a d-dimensional sphere of radius $\frac{3}{8}r$. On this flat surface, the surface area of the projection of $Q \cap P$ is therefore bounded by $A(\frac{3}{8}r, d)$. Projecting it back to the surface of S, we get that $A(Q \cap P) \leq 2^{d/4}A(\frac{3}{8}r, d)$. Putting it all together, we get that $G(Q) \leq k2^{1-3d/4}$. □

A.3 An Upper Bound on Privacy

We define a specific isolator I_2 and prove a lower bound on the probability with which it c-isolates an arbitrary RDB, i.e., c-isolates some point of the RDB. We begin by describing an isolator I_1 and prove a lower bound on the probability with which the "isolator" I_1 wins the following game with the "sanitizer".

First the sanitizer $\mathcal{S}$ chooses an arbitrary RDB R, chooses a point $x \in R$, and sanitizes R to obtain R'. Then the isolator is told x' and it performs a probabilistic algorithm on R' and x' to produce a point q. The isolator wins the game if q c-isolates x. Initially the sanitizer not only can choose the RDB R but also can choose a point $x \in R$ and challenge the isolator to c-isolate this point. Suppose we show that I_1 wins this game with probability at least p. That is, for each R and x chosen before the sanitization, $I_1(R', x')$ c-isolates x with probability at least p (over the random choices of I_1 and $\mathcal{S}$). Let I_2 be the isolator that, given only R', chooses a random point $x' \in R'$ and outputs $q = I_1(R', x')$. It follows that I_2 c-isolates R with probability at least p. Both adversaries, I_1 and I_2, use time polynomial in n.

As before, let $B(p, r)$ denote a ball of radius r around p.

Fact 13. *For all $x_1, x_2 \in \mathbb{R}^d$ and all $r_1, r_2 > 0$, if y is a point chosen uniformly at random from $B(x_2, r_2)$ then $\Pr[\, y \in B(x_1, r_1)\,] \le (r_1/r_2)^d$. If moreover $B(x_1, r_1) \subseteq B(x_2, r_2)$ then $\Pr[\, y \in B(x_1, r_1)\,] = (r_1/r_2)^d$.*

Recall that the point q *c-isolates* the point $x \in R$ if $B(q, c\|x - q\|)$ contains fewer than t points of R. It will be more convenient to work with approximation rather than isolation. The point q *ε-approximates* x if $\|x - q\| < \varepsilon \cdot \tau_{\mathrm{x}}$.

Lemma 14. *Let $c > 1$ and $x \in R$. If q c-isolates x then q $\frac{1}{c-1}$-approximates x.*

Proof. Let $\delta = \|x - q\|$. If q does not $1/(c-1)$-approximate x then $\delta \ge \tau_{\mathrm{x}}/(c-1)$. Let p be one of x's t nearest neighbors in R, so $\|x - p\| \le \tau_{\mathrm{x}} \le (c-1)\delta$. Since, by the triangle inequality, $\|q - p\| \le \|q - x\| + \|x - p\|$, we have $\|q - p\| \le \delta + (c-1)\delta = c\delta$. So $B(q, c\delta)$ contains at least $t + 1$ points of R. This contradicts that q c-isolates x. □

Lemma 15. *Let $c \ge 0$. If q $\frac{1}{c+1}$-approximates x then q c-isolates x.*

Proof. Let $\delta = \|x - q\|$, so $\delta < \tau_{\mathrm{x}}/(c+1)$. Let p be any point in $B(q, c\delta)$. By the triangle inequality, $\|p - x\| \le \|p - q\| + \|q - x\| \le c\delta + \delta = (c+1)\delta < \tau_{\mathrm{x}}$. Because at least one of x's t nearest neighbors has distance exactly τ_{x} from x, it follows that $B(q, c\delta)$ contains at most t points, *i.e.*, x and at most $t-1$ nearest neighbors of x. □

Let $\tau_{R'}(x')$ denote the t-radius of the sanitized point x' with respect to the sanitized database R', *i.e.*, the distance from x' to the t-th nearest neighbor of x' in R'.

The method used by the isolator is to compute $\tau_{R'}(x')$ (in time polynomial in n and d by the obvious algorithm) and use this to get an upper and lower bound on τ_{x}. The estimate of τ_{x} is then used to ε-approximate x (with some probability). The first lemma gives an upper bound on this probability assuming that some upper and lower bound on τ_{x} are known.

Lemma 16. *Let the sanitizer be* SSan *or* BSan. *There is a polynomial-time isolator I_0 such that, when given R', x', and positive numbers a, b, and ε such that $a \cdot \tau_{\mathrm{x}} \le \tau_{R'}(x') \le b \cdot \tau_{\mathrm{x}}$, the isolator ε-approximates x with probability at least*

$$\left(\frac{a}{b}\right)^d \left(\frac{\varepsilon}{\varepsilon + 1}\right)^d.$$

Proof. Let $r = \tau_{\mathrm{x}}$ and $r' = \tau_{R'}(x')$. The isolator I_0 first computes r'. It then chooses a point q uniformly at random from $B(x', (r'/a)(\varepsilon + 1))$, and outputs this point.

Now $\|x - x'\| \le r$ and $ar \le r'$ imply that $B(x, \varepsilon r) \subseteq B(x', \varepsilon r + r) \subseteq B(x', (r'/a)(\varepsilon + 1))$. Also, $r' \le br$ implies that the isolator chooses q from a ball of radius $(b/a)(\varepsilon + 1)r$. So the probability that q belongs to $B(x, \varepsilon r)$ is at least the radius εr divided by the radius $(b/a)(\varepsilon + 1)r$, all raised to the d-th power, *i.e.*, $(\varepsilon r/((b/a)(\varepsilon + 1)))^d$. □

It is not surprising that I_0's probability of isolation decreases as ε^d when ε decreases. A weakness of I_0 is that its probability of isolation decreases as $(a/b)^d$, so we should try to make the ratio a/b be as large as possible. But for the values of a and b shown below, a/b is "small". For the isolator I_1 given next, the ratio appears in the isolation probability as $1/\ln(b/a)$ so it not so important to make a/b small. The isolator I_1 uses I_0 as a subroutine.

Lemma 17. *Let the sanitizer be* SSan *or* BSan. *There is a polynomial-time isolator I_1 such that, when given R', x', and positive numbers a, b, and ε such that $a \cdot \tau_{\mathrm{x}} \le \tau_{R'}(x') \le b \cdot \tau_{\mathrm{x}}$, the isolator ε-approximates x with probability at least*

$$\frac{1}{ed\,\ln(b/a)} \left(\frac{\varepsilon}{\varepsilon + 1}\right)^d.$$

Proof. For some $\rho > 1$ chosen below, the interval $[a, b]$ is broken into intervals $I_i = [a\rho^i, a\rho^{i+1}]$ for $0 \leq i \leq m \stackrel{\text{def}}{=} \log_\rho(b/a) - 1$. The isolator I_2 chooses the interval i uniformly at random and then performs I_0 using $a' = a\rho^i$ and $b' = b\rho^{i+1}$ for I_0's input. The probability that I_2 ε-approximates x is the probability $1/(m+1)$ that $a' \cdot \tau_{\mathrm{x}} \leq \tau_{R'}(x') \leq b' \cdot \tau_{\mathrm{x}}$ times the probability that I_0 ε-approximates x, and the latter probability is multiplicative in $(a'/b')^d = \rho^{-d}$.

To choose ρ, the quantity to maximize is $Q(\rho) = \rho^{-d}/(m+1) = \rho^{-d}/\ln_\rho(b/a)$. This equals $(1/\ln(b/a))(\rho^{-d}\ln\rho)$. Setting its derivative to zero, it is maximized when $\rho = e^{1/d}$. Substituting this value of ρ gives $Q = 1/(ed\ \ln(b/a))$, which proves the lemma. ☐

The next two lemmas give values for a and b.

Lemma 18. *Let the sanitizer be* BSan. *Let $0 < p < 1$ and $a = (pt/n)^{1/d}$. For all RDBs R and all $x \in R$*

$$\Pr[\, a \cdot \tau_{\mathrm{x}} \leq \tau_{R'}(x')\,] \geq 1 - p.$$

Proof. Let $r = \tau_{\mathrm{x}}$ and $r' = \tau_{R'}(x')$. Suppose that all points of R except x have been sanitized. Let S be this set of $n-1$ sanitized points. If $r' < ar$ there must be at least t points s in $S - \{x'\}$ such that $x' \in B(s, ar)$. Let G be the set of points that belong to the intersection of at least t of the balls $B(s, ar)$ for $s \in S - \{x'\}$. Because there are a finite number $n-1$ of these balls, it is easy to see that $\mathsf{Vol}(G) \leq (n/t)B(ar)$. (There are a finite number of pairwise disjoint connected open sets W such that (i) the union of the W's equals G minus a set of measure zero, and (ii) for every W there is a subset of at least t of the balls $B(s, ar)$ such that every point of W belongs to exactly this subset of balls. Therefore, every component W of T "uses" at least $t \cdot \mathsf{Vol}(W)$ of the total volume $n \cdot B(ar)$ of the balls.)

Because x' is chosen randomly from a ball of volume $B(r)$,

$$\Pr[\, x' \in G\,] = \frac{\mathsf{Vol}(G)}{B(r)} \leq \frac{(n/t)B(ar)}{B(r)} = (n/t)a^d = p.$$

☐

Lemma 19. *Let the sanitizer be* BSan *or* SSan. *For all R and $x \in R$, $\tau_{R'}(x') \leq 4 \cdot \tau_{\mathrm{x}}$.*

Proof. Let $r = \tau_{\mathrm{x}}$. Obviously, $\|x - x'\| \leq r$. There are at least t points $y \in B(x, r) - \{x\}$, and $\tau(y) \leq 2r$ for each y, so $\|x - y\| \leq r$ and $\|y - y'\| \leq 2r$ for each y. By the triangle inequality, $\|x' - y'\| \leq 4r$. ☐

Combining Lemmas 15, 17, 18, and 19 gives lower bounds on the probability of c-isolation. For example, taking $p = 1/2$ and $\varepsilon = 1/(c+1)$ gives the following.

Theorem 20. *Let the sanitizer be* BSan. *There is a polynomial-time isolator such that, for all RDBs, the isolator isolates (some point of the RDB) with probability at least*

$$\frac{1}{2e(\ln(2n/t) + d\ln 4)}\left(\frac{1}{c+2}\right)^d.$$

A.4 A Positive Result for a Restricted Adversary

We give an upper bound on the probability that some point q, which is sufficiently near an SDB point x_2', can c-isolate an RDB point x_1. The definition of "sufficiently near" is relative to τ_{x_2}. As an immediate corollary, this is also an upper bound on the probability that an isolator can c-isolate an RDB R when the isolator is restricted to produce a point q that is close to a point in the SDB R'.

Theorem 21. *Let R be an arbitrary RDB and let n be the number of points in R. Let $0 < p \leq 1$, $b \geq 0$, and $c \geq 1$. If $n \leq \sqrt{p\left(\frac{c}{2+b+bc}\right)^d}$ and the isolator I must choose a point q within distance $b \cdot \tau_{\mathrm{x}_2}$ of some sanitized point x_2', then the probability that I isolates R is at most p.*

Proof. We show that

$$\Pr[\,(\exists\, x_1 \in R)\,(\exists\, x_2' \in R')\,(\exists\, q \text{ with } \|q - x_2'\| \le b \cdot \tau_{\mathrm{x}_2})\,[q \text{ } c\text{-isolates } x_1]\,] \le p \tag{6}$$

where the probability is over the random choices made in sanitizing R to produce R'. This immediately implies the theorem.

For (x_1, x_2) an arbitrary pair of points in R (possibly $x_1 = x_2$) and for all q with $\|q - x_2'\| = b \cdot \tau_{\mathrm{x}_2}$, we show that $\Pr[\,q$ c-isolates $x_1\,] \le p/n^2$. Because there are n^2 pairs (x_1, x_2), this proves (6). Let $r_2 = \tau_{\mathrm{x}_2}$ and $\delta = \|x_1 - q\|$. That q c-isolates x_1 implies that $B(q, c\delta)$ contains fewer than $t+1$ points of R. That x_2' is a sanitization of x_2 implies that $\|x_2 - x_2'\| \le r_2$, so $B(x_2', 2r_2)$ contains at least $t+1$ points of R. Therefore, $B(x_2', 2r_2)$ cannot be contained in $B(q, c\delta)$. Because $\|q - x_2'\| = br_2$ this implies that $2r_2 > c\delta - br_2$, so $\delta < (2+b)r_2/c$. So we must have $x_2' \in B(x_1, (2+b+bc)r_2/c)$. Because x_2' is randomly chosen from $B(x_2, r_2)$, the probability that $x_2' \in B(x_1, (2+b+bc)r_2/c)$ is at most $((2+b+bc)/c)^d \le p/n^2$. □

B Histogram Sanitizations

This section discusses how to prove that histogram-based sanitizations preserve privacy under our definition. Consider some partition of our space $\mathbb{R}^d$ into disjoint cells. The histogram for some dataset RDB is a list describing how many points from the dataset fall into each cell. A sanitization procedure is *histogram-based* if it first computes the histogram for the dataset and bases the output only on that information.

For example, in one dimension the cells would typically be sub-intervals of the line, and the histogram would describe how many numbers from the dataset fall into each of the intervals. In higher dimensions, the possibilities are more varied. The simplest partition divides space into cubes of some fixed side-length (say at most 1). That is, each cell is a cube $[a_1, a_1+1] \times \cdots \times [a_d, a_d+1]$ for integers $a_1, ..., a_d$.

We prove below that histogram sanitizations are private for several variants on the partition described above. We assume that the original data set RDB consists of n points drawn independently and uniformly from some large cube $[-1,1]^d$. Our goal is both to illustrate techniques that are useful for proving privacy, and to establish a result which we will need when we deal with cross-training-based sanitizers later on.

The main result of this section is the following:

Proposition 22. *Suppose that the adversary knows only that the dataset consists of n points drawn i.i.d. from the cube $[-1,1]^d$, and that we release the exact histogram (cell counts) for the natural partition of this cube into 2^d subcubes of side-length 1. Then the probability that the adversary succeeds at c-isolating a point for $c > 121$ is at most $2^{-\Omega(d)}$, as long as $t = 2^{o(d)}$.*

The constant 121 in the lemma can in fact be improved significantly, to approximately 30, with minor changes to the proofs in this section.

This result is strong—it essentially states that for any point q which the adversary might produce *after* seeing the histogram, the distance to q's nearest neighbor is at most a constant less than the distance between q and its $2^{\Omega(d)}$-th nearest neighbor. When $n = 2^{o(d)}$, the result is perhaps less surprising: the distance between q and its nearest neighbor is $\Omega(\sqrt{d})$ with high probability, and $2\sqrt{d}$ is an upper bound on the distance from q to its farthest neighbor (assuming q is in the large cube $[-1,1]^d$). For very large values of n (say $2^{\omega(d)}$), the proof becomes much more involved.

The next few subsections describe the proof of Proposition 22. Following that proof, we extend the proposition to the case where the cube is partitioned into a finer grid, containing $O(n2^d)$ equal-sized subcubes (Section B.2.1). Finally, in Section B.4, we consider a recursive partitioning of space into cells of varying sizes depending on the local density.

B.1 Simplifying the Adversary

We would like to understand the probability that the adversary isolates a point after seeing the sanitization, given reasonable assumptions about the adversary's a priori view of the database. We distinguish two related definitions of isolation. Recall that the t-radius of q with respect to a point-set D, denoted τ_{q}, is the distance from q to its t-th nearest neighbor in D.

The adversary is always given the sanitized database SDB as input (the adversary may also receive side information about the real database—typically, in our setting, the distribution from which the points in the real database are drawn).

- A **ball adversary** produces a pair (q, r) where $q \in \mathbb{R}^n$ is a point in space and $r \in \mathbb{R}^+$ is a non-negative real number. The adversary succeeds if $B(q, r)$, the ball of radius r centered at q, contains at least one point in RDB, but $B(q, cr)$ contains fewer than t points in RDB (equivalently, $r < \tau_{\mathrm{q}}/c$).
- A **point adversary** produces only a point $q \in \mathbb{R}^n$. The adversary succeeds at c-isolation if there is a point in D within distance τ_{q}/c of q, i.e. if there exists some r for which the corresponding ball adversary would have won.

We first prove Proposition 22 for ball adversaries since their behavior is easier to understand. At the end of the proof, we show that point adversaries do essentially no better than ball adversaries in this context, and so the restriction to ball adversaries is made without loss of generality.

B.2 Proving that Histograms Protect Privacy

Recall that in Proposition 22, the real data RDB is drawn uniformly from the large cube $[-1, 1]^d$, and the sanitization consists of the number of points from RDB contained in each of the cubes obtained by dividing $[-1, 1]^d$ once along each dimension. That is, a cell C is a d-dimensional hypercube of side-length and volume 1, which has one vertex at $(0, 0, ..., 0)$ and the opposite vertex in the set $\{-1, +1\}^d$. The total number of points in the database is n, and we denote the number of points appearing in a cell C by n_C. The sanitization is simply the list of all 2^d values n_C.

Define a function $f : \mathbb{R}^n \to \mathbb{R}^+$ which captures the adversary's view of the data.

$$f(x) = \frac{n_C}{n} \cdot \frac{1}{\mathsf{Vol}(C)} \quad \text{for } x \in C. \tag{7}$$

The function f is a probability density function. The adversary does not see the data as being drawn i.i.d. from according to f, but the function is useful nonetheless for bounding the adversary's probability of success.

Lemma 23. *If a ball adversary succeeds at c-isolation with probability ε, then there exists a pair (q, r) such that $\Pr_f[B(q, r)] \geq \varepsilon/n$ and $\Pr_f[B(q, cr)] \leq (2t + 8\log(1/\varepsilon))/n$.* [10]

The intuition for the proof of Lemma 23 is simple: it is sufficient to show that if one considers only the number of points landing in a particular region, there is almost no difference between the adversary's view of RDB and a set of n points sampled i.i.d. from f. A proof is given in Section B.3.

Roughly, Lemma 23 means that for a ball adversary to succeed, a necessary (but not sufficient) condition is that:

$$\frac{\Pr_f[B(q, r)]}{\Pr_f[B(q, cr)]} \geq \frac{\varepsilon/n}{(2t + 8\log(1/\varepsilon))/n} = \varepsilon/(2t + 8\log(1/\varepsilon)). \tag{8}$$

This means that it is sufficient to bound the ratio o the left-hand side above by some negligible quantity to prove privacy against ball adversaries (in fact, the upper bound need only hold as long as $\Pr_f[B(q, r)]$ is itself not too large). The better the bound on the ratio in Eqn. (8), the better the result on privacy. To get the parameters described in the statement of Proposition 22, it will be sufficient to prove a bound of $2^{-\Omega(n)}$ for the ratio: we can then think of ε as $2^{-\gamma d}$, for a constant $\gamma < 1$, and think of t as being $2^{-o(d)}$.

Our upper bound, and the proof of Proposition 22, rely on the following lemma:

Lemma 24. *There is a constant $1/60 < \beta < 1$ such that, for any point q, radius $r > 0$ and cell C for which $B(q, r) \cap C \neq 0$, we have:*

1. *If $r \leq \beta\sqrt{d}$, then*

$$\frac{\mathsf{Vol}(B(q, r) \cap C)}{\mathsf{Vol}(B(q, 3r) \cap C)} \leq 2^{-\Omega(d)}. \tag{9}$$

[10] We are assuming that n is at least $2t + 8\log(1/\varepsilon)$. The assumption is not necessary, but simplifies the proofs. In fact, when n is small one can use completely different proofs from the ones described here which are much simpler.

2. If $r > \beta\sqrt{d}$, then for all cells C' neighboring C:

$$C' \subseteq B(q, c_{\text{secure}} r) \tag{10}$$

where $c_{\text{secure}} \leq (2/\beta) + 1$. A neighboring cell is a cube of the same side-length which shares at least one vertex with C.

Remark: The true value of the constant β is actually much larger, at least 1/15. Obtaining this bound requires more careful versions of the proofs below.

We can now prove Proposition 22, which states that releasing the histogram preserves privacy, with the adversary's success probability bounded by $2^{-\Omega(d)}$. We first prove the lemma for ball adversaries, and then observe that (almost) the same proof works for point adversaries too.

Proof of Proposition 22: Ball adversaries: Assume there is a ball adversary who chooses the best possible pair (q, r) based on SDB.

First, suppose that $r \leq \beta\sqrt{d}$ (the constant is from Lemma 24). In that case, we will actually show that 3-isolation (not 121-isolation!) is possible only with very small probability. Our proof relies on Part 1 of the lemma. We can write the mass of $B(q, r)$ under f as a weighted sum of the volume of its intersections with all the possible cubes of C:

$$\Pr_f[B(q,r)] = \sum_C \frac{n_C}{n} \cdot \mathsf{Vol}(B(q,r) \cap C)$$

We can bound each of these intersections as an exponentially small fraction of the mass of $B(q, 3r)$:

$$\Pr_f[B(q,r)] \leq \sum_C \frac{n_C}{n} \cdot 2^{-\Omega(d)} \cdot \mathsf{Vol}(B(q,3r) \cap C) = 2^{-\Omega(d)} \cdot \Pr_f[B(q,3r)]$$

Now the mass of $B(q, 3r)$ is at most 1, which means that the ratio in Eqn. (8) is at most $2^{-\Omega(d)}$, and so $\varepsilon/(2t + 8\log(1/\varepsilon)) \leq 2^{-\Omega(d)}$. This is satisfied by $\varepsilon = 2^{-\Omega(d)}$ (for essentially the same constant in the Ω notation), and so in this case, 3-isolating the point q is possible only with probability $2^{-\Omega(d)}$.

Now consider the case where $r > \beta\sqrt{d}$. If $B(q, r)$ doesn't intersect any cells C, then we are done since the ball captures a point with probability 0. If there is a cell C which intersects $B(q, r)$, then, by Part 2 of Lemma 24, $B(q, c_{\text{secure}} r)$ (for $c_{\text{secure}} \leq (2/\beta) + 1$) contains all the cells C' which are neighbors to C, in particular all of $[-1, 1]^d$. (Recall that our points are initially uniform in $[-1, 1]^d$, and we consider a subdivision which splits the cube into 2^d axis-parallel subcubes of equal size). The adversary succeeds with probability zero at c_{secure}-isolation, since all n points will be within distance $c_{\text{secure}} r$.

Point adversaries: Suppose the adversary outputs a particular point q, and let r be the smallest radius such that $\Pr_f[B(q, r)] = \varepsilon$. By the previous discussion, $B(q, c_{\text{secure}} r)$ contains mass at least $(2t+8\log(1/\varepsilon))/n$. Thus, with probability at least $1-2\varepsilon$, there is no point inside $B(q, r)$ and there are t points inside $B(q, c_{\text{secure}} r)$ (by the proof of Lemma 23). The ratio between the t-th nearest point and the nearest point to q is then at most $c_{\text{secure}} r/r = c_{\text{secure}}$. The point adversary succeeds at c_{secure}-isolating a point with probability at most 2ε.

Because $\beta > 1/60$, the constant c_{secure} is at most 121, and so the adversary fails to 121-isolate any point in the database.

□

B.2.1 A Finer Subdivision

The proof of Proposition 22 extends easily to the case where the cube is divided into $\Omega(2^d n)$ subcubes. Imagine that points are drawn uniformly from a cube of volume 1 and we cut it into a fine grid of subcubes of volume v (note that $v = (2/k)^d$ for some integer k). We show that as long as v is large enough, releasing a histogram of the data over this fine grid preserves privacy with a high probability.

We begin by defining a bad event and bounding its probability.

Lemma 25. *Let $v = \Omega(\frac{t \log^2 n}{3^d n})$ and consider dividing the cube of volume 1 into subcubes of volume v each. The probability that there is a subcube with at least one point, surrounded by subcubes with no more than t points, is at most $n^{-\Omega(t \log n)}$.*

Proof. Consider any subcube C. The total volume of subcubes surrounding C is $3^d v$. The probability that C contains at least one point while the surrounding subcubes contain at most $t-1$ points is at most

$$v\binom{n}{t}(1-3^d v)^{n-t} \leq vn^t(1-3^d v)^{n-t+1}$$

There are v^{-1} choices for the central subcube C. Using

$$v = \Omega(\frac{t\log^2 n}{3^d n})$$

yields probability approximately $(1/n)^{\Omega(t\log n)}$. □

Note that the probability in the lemma is subpolynomially small in n, and it works for any n. In practice we only need a probability exponentially small in d. So if, say, $n = 2^d$, then it suffices to have $v > \Omega(\frac{t\log n}{3^d n}) = \Omega(\frac{2td}{6^d})$.

Having bounded the probability of the bad event, we show that if the bad event does not happen, then the probability of isolation is exponentially low.

Theorem 26. *Let c_{secure} be as in Lemma 24, let $v = \Omega(\frac{t\log^2 n}{3^d n})$ and consider dividing the cube of volume 1 into subcubes of volume v each. The probability, given a histogram over the subcubes, that an adversary c_{secure}-isolates a point in* RDB *is at most $2^{-\Omega(d)}$.*

Proof. Assume that the bad event described in Lemma 25 does not happen, and the ball adversary produces the guess (q, r). We can define a mass function f as in the proof of Proposition 22 (see Eqn. (7)).

As in the proof of Proposition 22, if r is small (less than a β fraction of the side-length of the cells), the ratio of the masses $\Pr_f[B(q,r)]/\Pr_f[B(q,3r)]$ is $2^{-\Omega(d)}$, and so the ratio $\varepsilon/(2t + 8\log(1/\varepsilon))$ can be bounded above by $2^{-\Omega(d)}$. If r is large and $B(q,r)$ intersects a particular cell C, then $B(q, c_{\mathrm{secure}}r)$ covers all 3^d neighboring cells C' of C. Since the bad event did not occur, if C contains any points at all then its neighborhood contains at least t points. In particular, the adversary will fail to c_{secure}-isolate any points. □

B.3 Proofs of Histogram-related Lemmas

Proof of Lemma 23: Because we are assuming that RDB consists of n points drawn i.i.d. uniformly from the cube $[-1,1]^d$, the distribution of the real database RDB *conditioned* on the sanitization SDB can be described precisely: in each cell, the n_C points that are present are chosen uniformly from within the cell. This means that for a particular output (q, r), a ball adversary can compute his probability of success exactly. Since we are considering computationally unbounded adversaries, we may assume without loss of generality that the adversary will choose the *best* possible output (q_0, r_0).

Let $N_{q,r}$ be the random variable representing the number of points that land in the ball $B(q,r)$ given the adversary's knowledge SDB. We can write $N_{q,r}$ as a sum of n *independent*, but not identically distributed, binary random variables. For every cell C, there are n_C variables in the sum, and each has expectation $\mathsf{Vol}(B(q,r)\cap C)/\mathsf{Vol}(C)$. The cells all have volume 1, so the expectation of $N_{q,r}$ is:

$$\mathbb{E}\left[N_{q,r}\right] = \sum_C n_C \frac{\mathsf{Vol}(B(q,r)\cap C)}{\mathsf{Vol}(C)} = n\cdot\Pr_f[B(q,r)]$$

Moreover, since $N_{q,r}$ is the sum of independent binary random variables, it is well-concentrated around its mean. If $\mu = n\cdot\Pr_f[B(q,r)]$, by a Chernoff bound (see, e.g., the textbook of Motwani and Raghavan):

$$\Pr[|N_{q,r} - \mu| \geq \delta\mu] \leq \exp(-\mu\delta^2/2) \tag{11}$$

All this means that even though the points in the database are not distributed i.i.d. according to f given SDB, from the point of view of the number of points captured by a given ball, they may as well be.

Returning to the adversary's best choice (q_0, r_0), there are two cases. First, suppose that the mass of $B(q_0, r_0)$ is at most ε/n. Then the expectation of N_{q_0,r_0} is at most ε, and so the probability that the ball captures even a single point is at most ε. This means that the adversary's success is bounded by ε, which is a contradiction to the hypothesis.

Second, suppose instead that $\Pr_f[B(q_0, cr_0)] \geq (2t + 8\log(1/\varepsilon))/n$. Then we can apply Eqn. (11) with $\delta = 1/2$ and $\mu \geq 2t + 8\log(1/\varepsilon)$. The probability that the ball captures fewer than $t \leq \mu/2$ points is at most

$\exp(-\mu/8) < \exp(-\log(1/\varepsilon)) = 2^{-(\ln 2)\log(1/\varepsilon)} < \varepsilon$, and so again the adversary fails with probability at least $1-\varepsilon$, which is a contradiction. $\square$

Proof of Lemma 24, Part 1:

We wish to prove that when the radius of a ball is at most $\beta\sqrt{d}$, doubling the radius of the ball increases the volume of its intersection with a cube C exponentially (in d). As with many proofs in this section, we will first prove the statement for the case $q \in C$, and then handle the case $q \notin C$ by a simple reduction.

The constant β itself arises from the following claim and its proof.

Claim 27. *Let C be a cell of side-length 1. There exists a constant $1/60 < \beta$ such that for any point q in C, the following holds: If $r < \beta\sqrt{d}$, then $\mathsf{Vol}(B(q,r)\cap C)/\mathsf{Vol}(B(q,2r)\cap C) \le 2^{-\Omega(d)}$.*

Proof of Claim:

For $x > 0$, define xC to be the cube $\{y \,|\, \forall i\, |Y_i| \le x\}$. So the side length of xC is $2x$.

We wish to prove that there exists a constant c such that for sufficiently large d, if B is a d-dimensional ball of radius $\sqrt{d}$ centered at any point $q \in cC$, and $2B$ is d-dimensional ball also centered at q, then $\mathrm{Vol}(2B\cap cC)/\mathrm{Vol}(B\cap cC)$ is exponential in d.

We write

$$\frac{\mathrm{Vol}(2B\cap cC)}{\mathrm{Vol}(B\cap cC)} = \frac{\mathsf{Vol}(2B)}{\mathsf{Vol}(B)}\cdot\frac{f(S(2B)\cap cC)}{f(B\cap cC)} \tag{12}$$

where $f(x\cap y)$ denotes the fraction of x that lies within y. Let $S(2B)$ denote the $(d-1)$-dimensional surface of the ball $2B$. Our strategy is to bound the ratio of intersections of spheres with the cube by a ratio of intersections of gaussians with the cube cC. This is because we can analyze the behavior of the gaussians one component at a time. That is, we can turn this into a question about simple one-dimensional normals.

Consider the ball $2B$. Imagine concentric shells, starting at q and increasing in radius until we reach the radius of $2B$. The fraction of each shell that resides inside cC decreases monotonically as the radius increases. Thus, using the fraction of the surface $S(2B)$ that lies within cC as a lower bound on the fraction of any such shell, we have

$$\mathrm{Vol}(2B\cap cC) > \mathsf{Vol}(2B)f(2B\cap cC)\,.$$

Let N denote a spherical gaussian centered at q with radius (standard deviation) $\sqrt{d}$. So N has the same radius as does B, and the ith coordinate of N is a chosen according to $\mathcal{N}(q_i,1)$, where $q=(q_1,\dots,q_d)$. For $k>0$, let kN denote a spherical gaussian centered at q with radius $k\sqrt{d}$, so each coordinate is chosen according to a a normal with variance k^2.

Consider $\frac{1}{2}N$, the spherical gaussian centered at q with radius equal to one half the radius of B. Note that most of the probability mass of $\frac{1}{2}N$ is inside B. Let us compare the distributions $(\frac{1}{2}N|B)$ (that is, sample $x\in\frac{1}{2}N$ and, if $x\notin B$, re-sample; repeat until $x\in B$) and B (that is, draw uniformly from B). In both cases, all the probability mass is inside B. However, the mass of $(\frac{1}{2}N|B)$ is more concentrated toward q and is very light toward the surface of B, while the mass of B is almost concentrated in a thin shell near the surface. Thus, following the same reasoning as above,

$$\mathrm{Vol}(B\cap cC) < \mathsf{Vol}(B)f((\frac{1}{2}N|B)\cap cC)\,.$$

We therefore have

$$\frac{\mathrm{Vol}(2B\cap cC)}{\mathsf{Vol}(B)} > \frac{\mathsf{Vol}(2B)}{\mathsf{Vol}(B)}\cdot\frac{f(2B\cap cC)}{f((\frac{1}{2}N|B)\cap cC)} \tag{13}$$

We now compute a lower bound on $f(2B\cap cC)$. Let $k>2$ be a parameter to be chosen later. Then

$$\Pr_{x\in kN}[x\in cC|x\notin 2B] < \Pr_{y\in S(2B)}[y\in cC] = f(2B\cap cC),$$

and so

$$f(2B\cap cC) \ge f(kN\cap cC) - f(kN\cap 2B) \tag{14}$$

Suppose $c>4k$. If we choose x_i from $\mathcal{N}(q_i,k^2)$ the probability that $q_i+x_i\in[-c,+c]$ is at least $1/4$. Thus, assuming $c>4k$, we have

$$f(kN\cap cC) > (1/4)^d. \tag{15}$$

In order to avoid worrying about the second term ($-f(kN \cap 2B)$), we will make k sufficiently large that this term is $o((1/4)^d)$.

We use a result of Indyk and Motwani [29], which holds for any $b > 0$:

$$\Pr_{x \in kN}[||x||^2 < k^2 d/b] < O(d)\exp(\frac{d}{2}(\frac{1}{b} - (1 - \ln b))) \tag{16}$$

Note that $k^2 d$ is the expected squared length of $x \in kN$. We will take $b = k^2/4$ in order to get a bound on the probability $\Pr_{x \in kN}[||x||^2 < 4d]$. This is because in order to be in $2B$, $||x||$ must be at most the radius of $2B$, which is $2\sqrt{d}$. So we get

$$\begin{aligned} f(kN \cap 2B) &= \Pr_{x \in kN}[||x||^2 < 4d] & (17)\\ &< O(d)\exp(-\frac{d}{2}(\frac{4}{k^2} - (1 - \ln(k^2/4)))) & (18) \end{aligned}$$

Rearranging this (starting by replacing $1 - \ln(\frac{k^2}{4})$ with $\ln \frac{e}{k^2/4}$), dropping the $4/k^2$ term, and taking $k > 8\sqrt{e}$, we get

$$f(kN \cap 2B) < O(d)(\frac{k}{2\sqrt{e}})^{-d} \in o(1/4)^d \tag{19}$$

Equations 14, 15, and 19 together imply

$$f(2B \cap cC) > f(kN \cap cC)/2 \tag{20}$$

We next give a simple upper bound on $f((\frac{1}{2}N|B) \cap cC)$ in terms of $f(\frac{1}{2}N \cap cC)$.

We imagine a procedure that repeatedly samples $\frac{1}{2}N$, computes the fraction of samples that lie in cC, and reports this fraction as an estimate of $f((\frac{1}{2}N|B) \cap cC)$.

On samples $x \in \frac{1}{2}N$ that lie outside B we are underestimating $f((\frac{1}{2}N|B) \cap cC)$ (because points inside B are more likeley to lie in cC). Since we are trying to compute an upper bound, we need to compensate for this underestimate. Fortunately, these "bad" samples are not numerous. For example, fewer than half of all samples from $\frac{1}{2}N$ will lie outside of B.

Viewed differently, we can imagine we are playing a game against an adversary. We are trying to prove that $f((\frac{1}{2}N|B) \cap cC)$ is small, and the adversary is trying to argue that it is large. We have a sampling procedure that on at most half the samples gives an answer the adversary might quibble with. To compensate, we allow the adversary to strike half of all the samples drawn. This affects the computed fraction of points lying inside cC, increasing the estimate by at most a factor of 2. This yields

$$f((\frac{1}{2}N|B) \cap cC) < 2f(\frac{1}{2}N \cap cC) \tag{21}$$

Plugging the bounds from Equations 20 and 21 into Equation 13 we obtain

$$\frac{\mathrm{Vol}(2B \cap cC)}{\mathrm{Vol}(B \cap cC)} > 2^d \cdot \frac{kN \cap cC}{4f(\frac{1}{2}N \cap cC)} \tag{22}$$

We will show that c can be made sufficiently large (as a function of k) so that the last term grows like $(\frac{1}{2} + \varepsilon)^d$, where ε is a positive constant.

At that point we will be done, since the right hand side of Equation 22 becomes $[2(\frac{1}{2} + \varepsilon)]^d = (1 + \delta)^d$ for some $\delta > 0$.

As noted earlier, we can now focus on a single dimension. We have q_i an arbitrary point in $[-c, c]$. Without loss of generality, assume $q_i < 0$. If we draw a normal distribution with mean q_i and any variance, the intersection of the interval $[-c, c]$ with the probability mass less than the mean (call this quantity b) is smaller than the intersection of the interval with the probability mass greater than the mean (call this larger quantity a). The distance to the interval boundary on the "a" side is at least c.

In order to compare the behavior of (now 1-dimensional) $\frac{1}{2}N$ (variance equal to 1/4) and kN (variance k^2), we consider a single curve and slide the endpoints of the interval in (for kN) and out (for $\frac{1}{2}N$). So we let a, respectively b, denote probability mass of the heavier (respectively, lighter) intersection with $[-c, c]$, and we let a' (respectively b') denote probability mass of the heavier (respectively, lighter) intersection with $[-c/2k, c/2k]$.

Our goal is to evaluate $(a'+b')/(a+b)$. First, note that $b' > b/2k$, because the mass is heaviest closest to q_i.

On the heavy side, the mass *lost* by moving the boundary in towards q_i is at most the original right tail, beyond $c/2k$. Using an upper bound on the tail [22], we get that

$$a' > \frac{1}{2} - \left(\frac{1}{c/2k}\right)\left(\frac{1}{\sqrt{2\pi}}\right)e^{-(c/2k)^2/2} \tag{23}$$

while, trivially, $a < 1/2$.

Taking $c \geq 2k(2\ln k)^{1/2}$ yields

$$\begin{aligned} a' &> \frac{1}{2} - \left(\frac{1}{(2\ln k)^{1/2}}\right)\left(\frac{1}{(2\pi)^{1/2}}\right)e^{-(2\ln k)/2} && (24)\\ &= \frac{1}{2} - \left(\frac{1}{(4\pi\ln k)^{1/2}}\right)\frac{1}{k} && (25)\end{aligned}$$

Since $k > 8\sqrt{e}$ and $a < \frac{1}{2}$, we get that

$$a' > \frac{1}{2} - \frac{1}{4k} > a(1 - \frac{1}{2k})$$

and so

$$a' + b' > (1 - \frac{1}{2k})a + \frac{1}{2k}b > \frac{1}{2}(a+b)$$

where the final inequality is strict, concealing an absolute constant.

The only assumptions on the constants in the previous proof were that $k > 8\sqrt{e}$ and $c \geq 2k(2\ln k)^{1/2}$. Taking $k = 13.1898$ and $c = 60$ satisfies these constraints. This completes the proof of the main subclaim, with a constant $\beta = 1/60$. □

To handle points q which are *not* in C, let q' be the nearest point in C to q. Any point which is in both $B(q,r)$ and C will also be in $B(q',r)$. In the other direction, $B(q,3r)$ contains $B(q,2r)$ as long as $B(q,r)$ intersects C at all. Hence

$$\frac{\mathsf{Vol}(B(q,r)\cap C)}{\mathsf{Vol}(B(q,3r)\cap C)} \leq \frac{\mathsf{Vol}(B(q',2r)\cap C)}{\mathsf{Vol}(B(q',2r)\cap C)}$$

Now $q' \in C$, and so we may apply the main subclaim to bound the right-hand side of this inequality by $2^{-\Omega(d)}$. (This completes the proof of Part 1 of Lemma 24) □

Proof of Lemma 24, Part 2: Suppose $B(q,r)\cap C \neq \emptyset$. The diameter of C is $\sqrt{d}$, so every point in C is at distance at most $\sqrt{d}+r$ from q. For any neighboring cell C', any point is at distance at most $2\sqrt{d}+r$ from C. Since $r > \beta\sqrt{d}$, we have

$$c_{\text{secure}}r = (2/\beta + 1)r > 2\sqrt{d} + r$$

Thus $C' \subseteq B(q, c_{\text{secure}}r)$, as desired. □

B.4 Recursive Refinement of Histograms

In this section we consider an extension of histogram sanitizations to variable sized cells. In particular, we assume, as before, that data is drawn uniformly at random from a d-dimensional cube. We divide the cube into 2^d equal-sized subcubes, as in Section B.2. Then, as long as there exists a subcube with at least $2t$ points from RDB, we further subdivide it into 2^d equal-sized cubes. We continue until all subcubes have fewer than $2t$ points, and release the exact number of points in each subcube.

The privacy of this sanitization is a simple extension of the results in Section B.2. We need the following analogue of Lemma 24. In the sequel, for a cell C, we use $P(C)$ to denote the "parent-cell" of C – this is the cell one level above C in the hierarchy, that was divided into 2^d parts including C.

Theorem 28. *Let c_{secure} be as in Lemma 24. The probability that an adversary, given a recursive histogram sanitization as described above can c_{secure}-isolate an* RDB *point is at most $2^{-\Omega(d)}$.*

Proof. As in the proof of Proposition 22, it is convenient to first prove the statement for *ball* adversaries. Suppose an adversary given SDB produces a pair (q, r). Consider the set of all cells in the recursive subdivision of $[-1, 1]^d$ which intersect $B(q, r)$. The side-length of these cubes varies,; we denote by ρ_C the side-length of a particular cell C.

First, suppose that there is some cell C intersecting $B(q, r)$ such that $r > \beta\rho_C\sqrt{d}$, where β is the constant from Lemma 24. Then by Part 2 of Lemma 24, $B(q, c_{\text{secure}} r)$ covers all of the parent cell $P(C)$. In particular, this larger ball captures at least $2t$ points, and so the adversary fails with probability 1.

Second, suppose that there is no cell C intersecting $B(q, r)$ for which the radius r is larger than $\beta\rho_C\sqrt{d}$. We can define a probability density function f as in the original proof of privacy of histograms (Section B.2, Eqn. (7)). Recall that the density f is constant within any particular cell C, and the mass it assigns to C is proportional to the number of points in C. By Part 1 of Lemma 24, the ball $B(q, 3r)$ captures an exponentially larger fraction of the probability mass under the function f. As long as $\varepsilon/(2t + 8\log(1/\varepsilon))$ is more than the fraction $\Pr_f[B(q, r)]/\Pr_f[B(q, 3r)]$, then either the small ball contains mass less than ε (and hence a point with probability at most ε), or with high probability the larger ball will capture at least t points, and again, the adversary will fail.

This proof can be converted to a proof for point adversaries as in the proof of Proposition 22. □

C Privacy for Cross-Training-Based Sanitizations

In this section, we combine a histogram-based sanitization with a spherical sanitization to obtain a provably private sanitization for $n = 2^{o(d)}$. We randomly divide the dataset into two sets — A and B. We construct a recursive histogram on B (as in Section B.4). We then sanitize points in A using only their position in the histogram on B. We release the sanitizations of points in A, along with the exact count, for every cell in the histogram, of points in A and B lying in that cell. We also assume that the adversary knows for every sanitized point $v' \in$ SDB, the cell in the histogram that its pre-image $v \in A$ lies in (this only helps the adversary).

For a point $v \in A$, let C be the cell containing v in the recursive histogram constructed for B. Let $P(C)$ be the parent cell of C. $P(C)$ has twice the side-length of C, and contains at least t points. Consider the following sanitization procedure:

> Let ρ_v be the side-length of the cube C. Select a point N_v at random from spherical Gaussian which has variance ρ_v^2 in each coordinate. Output $v' = v + N_v$.

As shown above in Section B.4, this procedure protects the privacy of points in B since they depend only on the recursive histogram for B. In this section we prove that it also protects the privacy of points in A, under the assumption that from the adversary's point of view, the *a priori* distribution of points in the database is uniform when restricted to the cell C.

Consider a particular point $v \in A$. Suppose the side-length of the cell C containing v is ρ_v. Lemma 29 shows that with probability $1 - 2^{-\Omega(d)}$ over the sanitized database SDB, the following occurs: for any point q which the adversary might produce, the distance between q and v will be $\Omega(\rho_v\sqrt{d})$ with probability $1 - 2^{-\Omega(d)}$. Since the t-radius of v is $O(\rho_v\sqrt{d})$, this is sufficient to show that adversary c-isolates v with probability at most $2^{-\Omega(d)}$ (for some constant c).

The result is quite useful. If A contains $2^{o(d)}$ points, then a union bound shows that with probability at least $1 - 2^{-(\Omega(d) - o(d))}$, the sanitization is "good": that is, the adversary can succeed at isolating some point with probability at most $2^{-\Omega(d)}$.

Remark 4: In fact, it seems likely that the union bound is not necessary here, and that one can use an argument similar to the proof of Proposition 22 to show that privacy holds even when the number of points is $n > 2^d$. There are quite a few difficulties however, and we do not yet have a complete proof of such a result. It is likely that the sanitization procedure will have to slightly modified, to either throw out or re-sanitize points v such that N_v is too large (say, when $\|N_v\| \geq 2\rho_v\sqrt{d}$).

Below, we state and prove the main lemma of this section:

Lemma 29. *Suppose that v is uniformly distributed in the cube C, and q is the output of the adversary on input $v' = v + N_v$. There is a constant $\alpha < 9$ such that with probability $1 - 2^{-\Omega(d)}$, the ball $B(q, \alpha||V - q||)$ contains the entire parent cell $P(C)$.*

Proof. The adversary's only information is the cube C over which v is distributed, and the point $v' = v + N_v$. Thus we can describe the posterior distribution on v exactly. It suffices to show that with high probability, this distribution has sufficient variance in every coordinate; it will follow that the distance to any point q will be $\Omega(\sqrt{d})$ with high probability.

We can consider the distributions of v and v' coordinate by coordinate, since the coordinates of both v and N_v are independent. Let U be uniformly distributed on $[-\frac{1}{2}, \frac{1}{2}]$, let $N \sim N(0,1)$ (that is, assume that the side-length of the cube C is $\rho_v = 1$). Conditioned on the event $W = U + N = w$, the random variable U is no longer uniform, but still reasonably well-spread. In fact, the conditional density of U given $W = w$ is proportional to the density of N at the point $w - u$:

$$f_{U|W=w}(u) = \frac{f_N(w-u)}{\Pr[N \in [-\frac{1}{2}, \frac{1}{2}]]}$$

The variance of U given $W = w$ is always non-zero, and so the expectation over w of the variance of U given $W = w$ is also strictly positive. Call this expectation $\bar{\sigma}^2$. It is tedious to compute exactly, but it is easy to approximate numerically using a package such as Maple.

$$\bar{\sigma}^2 = \mathbb{E}_w \left[\mathsf{Var}\left[U|W = w\right] \right] > 0.075 \tag{26}$$

We will assume for now that $q \in C$. There are three steps to the proof.

Step 1: For any fixed points v' and q, the expected square distance from v to q (conditioned on v') is a sum of *independent* random variables.

$$\|v - q\|^2 = \sum_{i=1}^{d} (v_i - q_i)^2 \tag{27}$$

The expectation of this sum is bounded below by the sum of the conditional variances in each coordinate. We use capital letters to distinguish random variables from particular values:

$$\mathbb{E}\left[\|V - q\|^2 | V' = v'\right] = \left(\sum_{i=1}^{d} \mathsf{Var}\left[U|W = v_i'\right]\right) + \left\|q - \mathbb{E}\left[V|v'\right]\right\|^2$$

To see why this holds, it is sufficient to expand each of the terms in the sum of Eqn. (27).

Step 2: Since we've assumed that $q \in C$, each term in the sum $\|V - q\|^2$ is bounded by 1. As long as the mean is large enough, we can apply the Chernoff bound to show that $\|V - q\|$ is concentrated around its mean. In particular, if the mean of $\|V - q\|^2$ is $\Omega(d)$, then the probability that $\|V - q\|^2$ deviates from its mean is at most $2^{-\Omega(d)}$ (see Eqn. (11)).

Step 3: Now we must show that with high probability over the choice of sanitization V', the expectation of $\|V - q\|^2$ is large for all q. We will again apply the Chernoff bound, but this time we will apply it to the sum of conditional variances $\sum_{i=1}^{d} \mathsf{Var}\left[U|W = v_i'\right]$. This sum is also a sum of independent random variables, each of which is bounded by 1 (however, the randomness here comes from the choice of V'). The mean of this sum is $d\bar{\sigma}^2$.

Putting these steps together: with probability $1 - 2^{\Omega(d)}$, the sum of conditional variance is at least $0.9 \cdot d\bar{\sigma}^2$. When this happens, the square distance $\|v - q\|^2$ is at least $(0.9)^2 d\bar{\sigma}^2$ with probability $1 - 2^{\Omega(d)}$. Finally, the t-radius of v will always be at most $2\sqrt{d}$, since that is the maxium diameter of the parent cell $P(C)$ when C has side length 1. Thus the ratio between the t-radius of v and the distance $\|v - q\|$ is less than $\alpha = 2/(0.9 \cdot \sqrt{\bar{\sigma}^2}) \leq 9$.

The case when q is outside of C reduces easily to the first case. Let q' be the point in C nearest to q. Once again, the distance $\|v - q'\|^2$ will be at least $r^2 = (0.9)^2 \cdot d\bar{\sigma}^2$ with high probability, and $B(q', \alpha r^*)$ contains $P(C)$. Since C is convex, the distance from q to v is at least the distance from to q' to v. This means that

$$B(q, \alpha\|v - q\|^2) \supseteq B(q', \alpha r^2) \supseteq P(C)$$

with high probability. □

The bound on the value of $\bar{\sigma}^2$ (Eqn. (26)) was obtained using the following straightforward script in Maple:

```
 # Density of a Gaussian with standard deviation 1.
npdf:= x -> exp(-x**2/2)/sqrt(2*Pi);
 # Density of W = N + U
fW:= w -> int(npdf(x),x=-w-(1/2)..-w+(1/2));
 # Conditional density of U given W=w
fUW:= (u,w) -> npdf(w-u) / fW(w);
 # Conditional expectation of U given W=w
expUW:= w -> int(u*fUW(u,w),u=-(1/2)..(1/2));
 # Variance of U condtioned on W=w
varUW:= w -> int( (u**2)*fUW(u,w),u=-(1/2)..(1/2)) - (expUW(w))**2;
 # Expected conditional variance of U given W=w, taken over w.
expvarUW := evalf(int(varUW(w)*fW(w), w=-infinity..infinity),4);
```

D Utility: Two Results on Clustering

D.1 Proofs for Learning a Mixture of Gaussians

Lemma 30 (Systematic Error). *Assume $d < n/2$. With probability at least $1 - (e^{-nw_{\min}/8} + 2^{-log^6 n/2})$, for all u*

$$|(I - P_{\overline{B}})A_u| \quad \leq \quad 16(\sigma_1 + \sigma_2)/w_u^{1/2}$$

Proof. Note that there are likely many columns v in $\overline{B}$ drawn from the same mean as u, ie columns v for which

$$(I - P_{\overline{B}})A_u \quad = \quad (I - P_{\overline{B}})B_v$$

With probability at least $1 - e^{-nw_u/8}$, column $(I - P_{\overline{B}})A_u$ occurs at least $nw_u/4$ times in $(I - P_{\overline{B}})B$. Let the row vector x have a 1 in each position that such a column exists. The definition of the L_2 norm (equation 1) requires

$$|(I - P_{\overline{B}})Bx|/|x| \quad \leq \quad \|(I - P_{\overline{B}})B\|_2$$

Note that $Bx = A_u|x|^2$, which leads us to

$$|(I - P_{\overline{B}})A_u| \quad \leq \quad \|(I - P_{\overline{B}})B\|_2/|x|$$

As with high probability $|x| > \sqrt{nw_u/4}$, we now work to bound the matrix norm on the right hand side. The triangle inequality implies that

$$\|(I - P_{\overline{B}})B\|_2 \quad \leq \quad \|(I - P_{\overline{B}})(B - \overline{B})\|_2 + \|\overline{B} - P_{\overline{B}}\overline{B}\|_2$$

$(I - P_{\overline{B}})$ is a projection matrix of unit L_2 norm, and so the first term is bounded by $\|\overline{B} - B\|_2$. The optimality of $P_{\overline{B}}\overline{B}$, as given in equation 2, bounds the second term by $\|\overline{B} - B\|_2$ as B is a rank k matrix. Combining these two bounds yields

$$\|(I - P_{\overline{B}})B\|_2 \quad \leq \quad 2\|B - \overline{B}\|_2$$

Finally, to bound $\|B - \overline{B}\|_2$ we use a result of Furedi and Komlos [26], which places a high probability bound on the norm of zero mean random matrices with independent entries. While $B - \overline{B}$ does not have independent entries – the sanitization process depends on the result of the random events producing $\widehat{B}$ – each of the matrices $B - \widehat{B}$ and $\widehat{B} - \overline{B}$ *do* meet the criteria for Furedi and Komlos's bound. As such, with probability at least $1 - 2^{-log^6 n/2}$

$$\begin{aligned}
\|B - \overline{B}\|_2 &\leq \|B - \widehat{B}\|_2 + \|\widehat{B} - \overline{B}\|_2 \\
&\leq 4\sigma_1\sqrt{n/2 + d} + 4\sigma_2\sqrt{n/2 + d} \\
&= 4(\sigma_1 + \sigma_2)\sqrt{n}
\end{aligned}$$

Combining these several bounds, we arrive at the statement of the theorem. □

Lemma 31 (Random Error). *For each u,*

$$Pr[|P_{\overline{B}}(A_u - \overline{A}_u)| > c(\sigma_1 + \sigma_2)\sqrt{k}] \quad \leq \quad 2ke^{-c^2}$$

We prove this lemma for the GSan sanitizer. However, the proof is easily modified to accommodate BSan and SSan sanitizers. Note that each are normalizations of GSan vectors, and will only change the projection by the factor of normalization. As we are discussing high dimensional Gaussians, the normalization factor will be highly concentrated, and unlikely to result in even a non-trivial constant factor.

Proof. We start by bounding

$$|P_{\overline{B}}(A_u - \overline{A}_u)| \quad \leq \quad |P_{\overline{B}}(A_u - \widehat{A}_u)| + |P_{\overline{B}}(\widehat{A}_u - \overline{A}_u)|$$

Both $A_u - \widehat{A}_u$ and $\widehat{A}_u - \overline{A}_u$ are d dimensional multivariate Gaussians with mean zero. These distributions are spherically symmetric, and their projection onto a fixed k dimensional subspace results in a k dimensional Gaussian in this space, with the same variance in each coordinate. The length of such vectors is described by the χ^2 distribution, and standard concentration bounds give the probability noted above. ☐

D.2 A Worst-Case Result for Arbitrary Distributions

A common data mining task is to do clustering on an RDB. Because only the SDB is available, an obvious approach is to run the clustering algorithm on the SDB and then map each cluster in it back to a cluster in SDB. Because there is a 1-1 correspondence between RDB points and SDB points, there is a natural way to do this mapping. The question, which we address in this section, is how good are the clusterings of RDB found by this approach.

For a dataset R, a *cluster C in R* is a nonempty subset of R. A *clustering $\mathcal{L}$ of R* is a partition $\{C_1, \ldots, C_k\}$ of R into clusters in R. We want to associate clusters in (resp., clusterings of) RDB with clusters in (resp., clusterings of) SDB. Because the sanitization procedures we consider in this paper sanitize each point of RDB independently, there is a natural way to do this.

Let R be an RDB and R' the corresponding SDB. If C is a cluster in R then $\sigma(C)$ is the cluster $\{\, x' \mid x \in C \,\}$ in R'. If $\mathcal{L} = \{C_1, \ldots, C_k\}$ is a clustering of R then $\sigma(\mathcal{L})$ is the clustering $\{\sigma(C_1), \ldots, \sigma(C_k)\}$ of R'. For C' a cluster in R' (resp., $\mathcal{L}'$ a clustering of R'), the cluster $\sigma^{-1}(C')$ in R (resp., the clustering $\sigma^{-1}(\mathcal{L}')$ of R) is defined analogously.

A *clustering (optimization) problem* is defined by specifying: (1) a *cost function* that maps each clustering in $\mathcal{L}$ to a real number; and (2) a set $F(R)$ of *feasible clusterings* of R. Given R, the objective is to find a clustering $\mathcal{L} \in F(R)$ having smallest *cost*. Let *opt*(R) be the minimum of *cost*$(\mathcal{L})$ over all $\mathcal{L} \in F(R)$. We consider the cost function *diam*$(\mathcal{L})$ = the maximum diameter of $C_1, \ldots, C_k$.

The next lemma gives a worst-case (over the cluster and the sanitization) upper bound on on the diameter of a sanitized cluster in terms of the diameter of the original cluster. This gives us hope that if we run a clustering algorithm on R' then it will not overlook a cluster C in R because the diameter of C exploded during sanitization. An exception is when the size of C is smaller than the crowd size t. In fact, in this case we *want* the sanitization to spread out the cluster.

A cluster C is *t-large* if $|C| \geq t$. A clustering $\mathcal{L}$ is *t-large* if all the clusters in $\mathcal{L}$ are t-large.

Fact 32. *If the cluster C is t-large and $x \in C$, then $\tau(x) \leq diam(C)$.*

Lemma 33. *Let R be an arbitrary RDB, let R' be a possible sanitization of R, and let $\mathcal{L}$ be a t-large clustering of R. Then*

$$diam(\sigma(\mathcal{L})) \leq 3 \cdot diam(\mathcal{L}).$$

Proof. Let x'_1, x'_2 belong to the same cluster C' of $\sigma(\mathcal{L})$. Let $x_i = \sigma^{-1}(x'_i)$ and $C = \sigma^{-1}(C')$. Then by Fact 32

$$\|x'_1 - x'_2\| \leq \|x_1 - x_2\| + \tau(x_1) + \tau(x_2) \leq 3 \cdot diam(C).$$

☐

In addition to the diameters of clusters in R becoming too large, we must also bound the amount by which they become too small. Otherwise, a random set S of points of R might sanitize to a good (small-diameter) cluster in R' and thus be mistakenly identified as a good cluster in R. Obviously, it is possible that this can happen, even when S is t-large, so we cannot prove an upper bound on the amount that a cluster can shrink that is worst-case over the cluster and the sanitization. However, a result can be shown when the diameter of a clustering is compared to the optimal diameter. The optimization is done over t-large clusterings containing at most k clusters, where t and k are fixed; call these the (t,k)-clusterings. Denote these optima by $opt_{t,k}(R)$ and $opt_{t,k}(R')$. In particular, we show: an algorithm that approximates the optimal diameter of a clustering of R' to within a constant factor gives an algorithm that constant-factor approximates the optimal diameter of a clustering of R.

Theorem 34. *Fix some t and k. Let R' be a possible sanitization of R and let $\mathcal{L}'$ be an arbitrary (t,k)-clustering of $R' = \sigma(R)$ such that $diam(\mathcal{L}') \leq r \cdot opt_{t,k}(R')$. Let $\mathcal{L} = \sigma^{-1}(\mathcal{L}')$ (so $\mathcal{L}$ is also a (t,k)-clustering). Then*

$$diam(\mathcal{L}) \leq (3r+2) \cdot opt_{t,k}(R).$$

Proof. Let x_1 and x_2 be two points that belong to the same cluster C in $\mathcal{L}$. Let $x_i' = \sigma(x_i)$, so x_1', x_2' belong to the same cluster $\sigma(C)$ of $\mathcal{L}'$. We show that $\|x_1 - x_2\| \leq (3r+2) \cdot opt_{t,k}(R)$. Let $\mathcal{L}^*$ be an optimal (t,k)-clustering of R, i.e., $diam(\mathcal{L}^*) = opt_{t,k}(R)$. Because $\sigma(\mathcal{L}^*)$ is a (t,k)-clustering of R', we have $diam(\sigma(\mathcal{L}^*)) \geq opt_{t,k}(R')$. Using this with Lemma 33 and the assumption that $diam(\mathcal{L}') \leq r \cdot opt_{t,k}(R')$,

$$\|x_1' - x_2'\| \leq r \cdot opt_{t,k}(R') \leq r \cdot diam(\sigma(\mathcal{L}^*)) \leq 3r \cdot diam(\mathcal{L}^*).$$

For $i = 1, 2$, x_i belongs to some t-large cluster C_i^* of $\mathcal{L}^*$. So Fact 32 gives $\tau(x_i) \leq diam(C_i^*) \leq diam(\mathcal{L}^*)$ for $i = 1, 2$. Therefore,

$$\begin{aligned} \|x_1 - x_2\| &\leq \|x_1' - x_2'\| + \tau(x_1) + \tau(x_2) \\ &\leq 3r \cdot diam(\mathcal{L}^*) + 2 \cdot diam(\mathcal{L}^*) \\ &= (3r+2) \cdot opt_{t,k}(R). \end{aligned}$$

□