Privacy Buckets: A numeric method for k-fold tight differential privacy

Sebastian Meiser[1], Esfandiar Mohammadi[2*]
[1] University College London, United Kingdom, e-mail: s.meiser@ucl.ac.uk
[2] ETH Zurich, Switzerland, e-mail: mohammadi@inf.ethz.ch

October 19, 2017

Abstract

The robustness of (approximate) differential privacy (DP) guarantees in the presence of thousands and even hundreds of thousands observations is crucial for many realistic application scenarios, such as anonymous communication systems, privacy-enhancing DB queries, or privacy-enhancing ML methods. Composition theorems capture DP under repeated observations, but previous work provides untight bounds, which can tremendously amplify after hundreds of thousands of compositions.

This work improves on previous work by providing upper and lower bounds for approximate DP, which enables us to quantify how untight our upper bound is. We introduce a numerical method and an implementation for computing provable upper and lower bounds for approximate DP for a given number of observations. In contrast to previous work, our bucketing method retains the shape of the distributions which enables us to compute tighter bounds. We show that, while previous work seems to be tight for the Laplace mechanism on statistical queries, our work is significantly tighter for other scenarios, such as the Gaussian mechanism on statistical queries or for real-world timing leakage data. We show that it is worth to conduct a tight privacy analysis by improving, as a case study, the privacy analysis of the anonymous communication system Vuvuzela. We show that for the same privacy target as in the original Vuvuzela paper, 5 to 10 times (depending on the assumed number of observations) less noise already suffices, which significantly reduces Vuvuzela's overall bandwidth requirement.

*both authors equally contributed to this work

Contents

1 Introduction

Privacy analyses of privacy-enhancing systems, such as anonymous communication systems [15], privacy-enhancing DB queries [4], and privacy-enhancing ML methods [1], play a crucial role in understanding the effectiveness of these systems. The notion of *differential privacy*, written as ε-DP [4] and its important relaxation *approximate differential privacy*, written as (ε, δ)-DP [13, 6] quantify, in terms of two parameters ε and δ, the privacy leakage against a strong worst-case attacker and have become a de-facto standard in the field. In many application scenarios, the privacy has to hold under continued observations, which enables potential attackers to make thousands if not hundreds of thousands of observations. The parameters ε and δ inevitably grow under continued observation and with them the degree of privacy deteriorates under a sufficiently long period of continued observation.

In the original (ε, δ)-DP paper [4, 13], a notion of privacy under continued observation (called k-fold (ε, δ)-DP) was presented and generic bounds were proven in a composition theorem. It turned out that these bounds could be dramatically improved. As understanding the precise degree of privacy under continued observations is crucial for understanding a system's realistic potential for enhancing privacy, there have been significant efforts [7, 10] to find tighter bounds on DP under continued observation.

While previous work [10] made significant improvements on the initial naïve bounds and even seems to be tight w.r.t. the Laplace mechanism, this paper shows that for other mechanisms these generic results are not tight and substantially better bounds can be found, e.g., for the Gaussian mechanism or for estimating the privacy leakage of timing side-channels. The previous work lacks one major ingredient: it is completely oblivious to the underlying distributions and cannot, by its very nature, make use of the nature of the distribution-specific behavior under continued observation. In a sense, such generic results inherently assume those distributions that leak most under continued observation.

1.1 Contribution

This work presents a numerical method and an implementation for computing provable upper and lower bounds for k-fold (ε, δ)-DP, for a given number of k observations. We show that our results are significantly tighter than the bounds from previous work [10]. As a short disclaimer, this work is meant to be a powerful tool for finding tight bounds in a (ε, δ)-DP privacy analysis of a privacy-enhancing system; this work does not attempt to contribute to the task of finding the right level of abstraction, a useful attacker model, suitable usage behaviors, the right privacy mechanism (e.g., the shape of the noise), and other tasks that are needed in a useful privacy analysis.

The notion of (ε, δ)-DP ultimately describes a relationship between a pair of distributions, such as the outputs of a privacy-enhancing mechanism on related inputs.[1] The core idea of this work is to compute for all atomic events x within a pair of distributions V, W the quotient $V(x)/W(x)$ and to throw all atomic events with the same quotient into the same bucket. This representation soundly abstracts away from the exact events within the distributions but retains the shape of the distributions. These buckets can be used to compute (ε, δ)-DP and the increase of ε and δ for a given number of observations. As there are in practice too many such buckets (particularly for distributions with an infinite support), we approximate these buckets by dividing the set of buckets into B sets, thereby approximating the buckets. We show that we can use these approximate buckets to compute (ε, δ)-DP under a up to a million observations. Our method can generically deal with distributions with finite support and we show how to embed the continuous Laplace distribution into these buckets, which does not have a finite support.

We illustrate that our method is tighter than previous work. We study how our bucketing approach compares to Kairouz et al.'s composition theorem [10]. While it seems as if Kairouz et al.'s composition theorem provides bounds that are tight for the Laplace mechanism, we show that they are not tight for the Gaussian mechanism or for a model of timing-leakage measurements from the anonymous communication system CoverUp [14]. We show that our bucketing approach is significantly tighter in these cases.

Moreover, our approach allows for analyzing how different distributions compose in terms of differential privacy. We find that the (ε, δ)-DP of Laplace noise under composition converges to the (ε, δ)-DP of Gauss noise under composition, if the Gaussian noise has half the variance of the Laplace noise. Note that this is <u>not</u> an example of the central limit theorem [2] as the composition is not the convolution but the product of distributions. Additionally, we show that for the same variance, Gaussian noise provides significantly stronger privacy guarantees under a high number of observations.

Finally, we apply our results to illustrate that a tight analysis of a privacy-enhancing system can lead to a significant reduction in the protocols overhead without reducing the required degree of privacy. As a case study, we improve the privacy analysis of the anonymity network Vuvuzela[15], which uses random noise to increase the privacy. First, we show that an improved analysis alone can enable a 2 to 4-fold reduction in noise while achieving the same privacy goals under hundreds of thousands of observations, depending on the number of observations.[3] If we do not reduce the amount of noise but keep the amount recommended in the Vuvuzela paper, we show that a precise analysis leads to privacy bounds that are roughly 3 to 4 orders of magnitude better. Second, we additionally propose to improve the shape of the noise from Laplace noise to Gaussian noise. In this case, we achieve a 5 to 10-fold reduction of noise for the same privacy goals, and if we stick to the amount of noise from the Vuvuzela paper, we show a 4 to 6 orders of magnitude improvements of the privacy bounds.

[1] As an example, when applying the Laplace mechanism to two databases with sensitivity 1, $(\varepsilon, 0)$-DP analyzes a pair of Laplace distributions with scale parameter $1/\varepsilon$ with mean 0 and mean 1.

[2] The central limit theorem shows that the sum of many independent random variables is normally distributed.

[3] The more observations are estimated the higher the error of a loose bound; hence, in those cases the tightness of our bounds leads to a more significant improvement.

The example of Vuvuzela highlights several important contributions of our approach for practical privacy-enhancing mechanisms: First, our bucketing method allows for a fast, uncomplicated (re-)evaluation of existing privacy analyses. Such a re-evaluation using state-of-the-art composition results such as Kairouz et al.'s composition theorem or our bucketing can yield impressively better results than naïve privacy bounds. Second, in contrast to Kairouz et al.'s generic composition theorem, our bucketing method retains the shape of the distributions which allows us to effectively compare different noise mechanisms and this can again significantly impact the resulting bounds. Third, our bucketing provides lower bounds and thus shows exactly to which extend our results could potentially be further improved. In many cases where the lower bounds (almost) equal the upper bounds our method is provably optimal (up to the slight difference in the bounds).

While our result expects concrete distributions as input, we show that in many cases concrete worst-case distributions can be found for k-fold (ε, δ)-DP with a given sensitivity. Worst-case distributions in this sense are the output distribution of the mechanism under worst-case inputs. As an example consider counting queries under the Gaussian mechanism on a database. While the definition of k-fold (ε, δ)-DP allows the attacker to choose two new databases that have a given sensitivity in every rounds (i.e., for every observation), it suffices to analyze in every round the leakage of a pair of the same Gaussian distributions with the same scale parameter and with means differing by the sensitivity of the databases.

The dominant factor in the runtime complexity of calculating a composition with our method is in the order of $O(B^2)$, where B is the number of buckets, i.e., the granularity of the approximation. If the worst-case inputs don't change from one observation to the next, which is the case for most applications of DP, we perform repeated squaring and thus only need to compute $\log_2(k)$ composition operations.

2 Related work

Differential privacy Differential Privacy (DP) [4] quantifies how closely related two similar distributions are from an information-theoretic perspective. The probability of any event in any one of the two distributions is almost the same as the probability of the event in the other distribution, bounded by a multiplicative factor e^{ε}, where ε is a small positive number and we say the distributions are ε-DP. Formally, we say that two distributions A and B over the universe $\mathcal{U}$ are ε-DP, if $\forall S \subseteq \mathcal{U}.\ \Pr[x \in S | x \leftarrow A] \leq e^{\varepsilon} \cdot \Pr[x \in S | x \leftarrow B]$ (and vice versa). To extend the applicability of DP, approximate differential privacy [5] allows for distributions to exceed limiting factor ε, as long as it is exceeded, in total only by a small value δ. Formally, we say that two distributions A and B over the universe $\mathcal{U}$ are (ε, δ)-DP, if $\forall S \subseteq \mathcal{U}.\ \Pr[x \in S | x \leftarrow A] \leq e^{\varepsilon} \cdot \Pr[x \in S | x \leftarrow B] + \delta$ (and vice versa). The notion of *computational differential privacy* [13] replaces the sets S of events by adversarial distinguisher machines.

How to use distributions to calculate differential privacy Classically, differential privacy is defined for all pairs of neighboring databases while relaxations leave the choice over the databases to an adversary. In either case, the notion argues about all possible such scenarios and adversarial choices, which is in contrast to our numerical approach: we require two concrete distributions, not a set of possible distributions. However, in practice, there is a direct connection between the worst-case choices of scenarios or adversarial decisions and very simple concrete distributions. For example, when considering sum queries with sensitivity 1 to which Laplacian noise is added, we can simply compare the respective Laplacian distributions with means 0 and 1 respectively instead of considering all possible combinations of neighboring databases. To formally apply our approach, we require the choice of two fitting distributions and the existence of a reduction: given a description of the scenario or an adversarial choice as well as an output of the distributions we consider in our calculation, the reduction produces the respective output within the differential privacy scenario of choice. Returning to our example of sum queries q for two neighboring databases D_1 and D_2 where the true answers for the databases are $q(D_1) = x$ and $q(D_2) = x + 1$ respectively, one can map any output y from a distribution to $y + x$ to obtain the correct adversarial view for the respective scenario.

Composition of differential privacy One of the main and most important benefits of differential privacy is that it withstands composition, albeit with a loss. The composition of ε-DP is straight forward: two distribution pairs that are respectively ε_1-DP and ε_2-DP together are $(\varepsilon_1 + \varepsilon_2)$-DP and, as long as ε is tight, this is the best possible result. If the distribution pairs are $(\varepsilon_1, \delta_1)$-DP and $(\varepsilon_2, \delta_2)$-DP, a straight-forward

result shows that their composition is $(\varepsilon_1+\varepsilon_2, \delta_1+\delta_2)$-DP, however, this result is not necessarily tight. Since many privacy-preserving systems are subject to small privacy leakages for every adversarial observation, previous work has investigated tighter bounds of (ε, δ)-DP under composition [7, 10], which we discuss in detail in Section 3.3. Other composition results focus on special cases and provide even tighter bounds [9, 12].

All these works have in common that they are oblivious to the actual distributions and their bounds only rely on the initial values for ε and δ.

Dependencies The work of Liu, Chakraborty and Mittal [12] discusses the importance of correctly measuring the sensitivity of databases for differential privacy. They show that in real-world examples entries can be correlated and thus cannot be independently exchanged as by DP's basic definition. Their approach, however, finally results in the same techniques being used to achieve the same goal: noise applied to database queries results in differential privacy, although the sensitivity is calculated in a more complex manner. Our results can directly be applied in such a setting as well: given the (final) distributions that potentially consider dependent entries we calculate differential privacy guarantees for these distributions.

Optimal distributions There has been recent work [8, 11] on finding optimal mechanisms for differential privacy for a large utility functions. However, these results concentrate on a single observation and do not characterize how these mechanism behave under k-fold composition.

3 Bucketing two distributions

3.1 Informal description of bucketing

Calculating differential privacy guarantees under composition is typically done independently of the shape of the underlying distributions, simply based on the differential privacy guarantees before the composition. This obliviousness is both the greatest strength and greatest weakness of this method: One doesn't need to know the shape of the distributions to give sound differential privacy guarantees under manifold composition.

We now introduce an alternative approach: We approximate the distributions and the ways they are related. Given two distributions A and B, for differential privacy the most important aspect of each event x is the factor between the probability that the event occurs in A, denoted by $\Pr[x \leftarrow A]$, and the probability that the same event occurs in B, denoted $\Pr[x \leftarrow B]$. Consequently, we group events by this factor $\frac{\Pr[x\leftarrow A]}{\Pr[x\leftarrow B]}$ and accumulate them with a very similar factor.

As our main aim is to compose the distributions efficiently and without unnecessary losses, we scale each group of events, which we also call a *bucket*, exponentially: given a factor $f > 1$, a bucket $\mathcal{B}(i)$ contains all events where $f^{i-1} < \frac{\Pr[x\leftarrow A]}{\Pr[x\leftarrow B]} \leq f^i$. The value $\mathcal{B}(i)$ is then the sum over the probabilities of all those events (according to distribution A). Thus, under composition of A and B with distributions C and D we can then simply combine buckets $\mathcal{B}(i)$ and $\mathcal{B}_C(j)$ multiplicatively and yield probability of all events in $\mathcal{B}_{A\times C}(i+j)$, for which we still have $\frac{\Pr[x\leftarrow A]}{\Pr[x\leftarrow B]} \cdot \frac{\Pr[x\leftarrow C]}{\Pr[x\leftarrow D]} \leq f^{i+j}$.

3.2 Differential privacy

We repeat the definition for approximate differential privacy and adapt it to suit our cause. Approximate differential privacy characterizes privacy by a multiplicative value ε and an additive error value δ. In particular, we introduce *tight differential privacy* to characterize the smallest values of ε and δ for which differential privacy is satisfied.

Definition 1 ((Tight) (ε, δ)-differential privacy)**.** *Two distributions A and B over the universe $\mathcal{U}$ are (ε, δ)-differentially private, if for every set S,*

$$\Pr[x \in S : x \leftarrow A] \leq e^{\varepsilon}\Pr[x \in S : x \leftarrow B] + \delta \text{ and}$$
$$\Pr[x \in S : x \leftarrow B] \leq e^{\varepsilon}\Pr[x \in S : x \leftarrow B] + \delta.$$

We say that A and B are tightly (ε, δ)-differentially private if they are (ε, δ)-differentially private, but $\forall \delta' < \delta$, A and B are not (ε, δ')-differentially private.

Note that we can also say that A and B are ε-tightly (ε, δ)-differentially private if they are (ε, δ)-differentially private, but $\forall \varepsilon' < \varepsilon$, A and B are not (ε', δ)-differentially private. However, we will consider ε to be a goal or input to our system and thus not pursue ε-tightness.

We argue that this can be characterized precisely by the following calculation:

Lemma 1. *For every ε, two distributions A and B over the finite universe U are tightly (ε, δ)-differentially private with*

$$\delta = \max\left(\sum_{x \in U} \max\left(\Pr\left[x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \leftarrow B\right], 0\right), \sum_{x \in U} \max\left(\Pr\left[x \leftarrow B\right] - e^{\varepsilon}\Pr\left[x \leftarrow A\right], 0\right)\right)$$

Proof. Let $\varepsilon \geq 0$ and let A and B be two distributions over the universe $\mathcal{U}$. We show the equivalence by first showing that (1) for every set S, the calculation describes an upper bound and then that (2) there exists a set S such that this bound is tight.

(1) We show that $\forall S \subseteq \mathcal{U}$,

$$\begin{aligned} &\Pr\left[x \in S : x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \in S : x \leftarrow B\right] \\ \leq &\sum_{x \in U} \max\left(\Pr\left[x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \leftarrow B\right], 0\right) \end{aligned}$$

The inverse direction then follows analogously.

$$\begin{aligned} &\Pr\left[x \in S : x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \in S : x \leftarrow B\right] \\ = &\sum_{x \in S} \Pr\left[x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \leftarrow B\right] \\ \leq &\sum_{x \in S} \max\left(\Pr\left[x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \leftarrow B\right], 0\right) \\ \leq &\sum_{x \in U} \max\left(\Pr\left[x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \leftarrow B\right], 0\right) \end{aligned}$$

(2) Let $S := \{x \in \mathcal{U} \text{ s.t. } \Pr\left[x \in A\right] \geq e^{\varepsilon}\Pr\left[x \in B\right]\}$. Then,

$$\begin{aligned} &\Pr\left[x \in S : x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \in S : x \leftarrow B\right] \\ = &\sum_{x \in S} \Pr\left[x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \leftarrow B\right] \\ = &\sum_{x \in U} \max\left(\Pr\left[x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \leftarrow B\right], 0\right). \end{aligned}$$

Analogously, for $S := \{x \in \mathcal{U} \text{ s.t. } \Pr\left[x \in B\right] \geq e^{\varepsilon}\Pr\left[x \in A\right]\}$,

$$\begin{aligned} &\Pr\left[x \in S : x \leftarrow B\right] - e^{\varepsilon}\Pr\left[x \in S : x \leftarrow A\right] \\ = &\sum_{x \in S} \Pr\left[x \leftarrow B\right] - e^{\varepsilon}\Pr\left[x \leftarrow A\right] \\ = &\sum_{x \in U} \max\left(\Pr\left[x \leftarrow B\right] - e^{\varepsilon}\Pr\left[x \leftarrow A\right], 0\right). \end{aligned}$$

Thus, for every pair of distributions A and B and for every $\varepsilon \geq 0$ the distributions are tightly (ε, δ)-differentially private, where δ is calculated as described.

□

Trade-off between ε and δ. Leveraging the calculations from above, we can immediately see that for every value of ε, there is an optimal value for δ, such that tight (ε, δ)-differential privacy holds. We can draw a graph portraying the relationship between those two variables. For some distributions there is a value ε_0, the distributions are ε_0-differentially private. However, (ε, δ)-differential privacy guarantees for arbitrarily small values of ε can be achieved as well, where δ is calculated as in Lemma 1.

3.3 Composition of differential privacy

One of the main advantages of differential privacy is the fact that guarantees are still sound under composition, albeit with increasing values for ε and δ.

Definition 2 (k-fold (ε, δ)-DP of a mechanism)**.** *A randomized algorithm M with domain $\mathcal{D}$ and range $\mathcal{U}$ is k-fold (ε, δ)-differentially private for sensitivity s if for all $S \subseteq \mathcal{U}^k$ and for all $(x_1, \ldots, x_k), (y_1, \ldots, y_k) \in \mathcal{D}^k$ such that $\forall 1 \leq i \leq k.\ ||x_i - y_i||_1 \leq s$:*

$$\Pr[(M(x_1), \ldots, M(x_k) \in S] \leq e^{\varepsilon} \Pr[M(y_1, \ldots, y_k) \in S] + \delta$$

Note that when we describe differential privacy in terms of distributions over the worst-case inputs, the composition of differential privacy is equivalent to considering differential privacy for product distributions. If x_0, x_1 are the worst-case inputs for a mechanism M, resulting in the distributions $M(x_0)$ and $M(x_1)$, then the k-fold composition is described Definition 1 on the distributions $A = M(x_0)^k$ and $B = M(x_1)^k$. Similarly, a composition of two different mechanisms M and M' with worst-case inputs x_0, x_1 and x'_0, x'_1 respectively, boils down to Definition 1 on the distributions $A = M(x_0) \times M'(x'_0)$ and $B = M(x_1) \times M'(x'_1)$.

The main composition results we compare our work with are: naive composition, slightly less naive composition and two composition result with improved bounds [7, 10]. We recall these results here.

Lemma 2 (Naïve Composition)**.** *Let (A_1, B_1) and (A_2, B_2) be two pairs of distributions, such that A_1 and B_1 are $(\varepsilon_1, \delta_1)$-differentially private and A_2 and B_2 are $(\varepsilon_2, \delta_2)$-differentially private. Then $A_1 \times A_2$ and $B_1 \times B_2$ are $(\varepsilon_1 + \varepsilon_2, \delta_1 + \delta_2)$-differentially private.*

Lemma 3 (Adaptive Composition)**.** *Let (A_1, B_1) and (A_2, B_2) be two pairs of distributions, such that A_1 and B_1 are $(\varepsilon_1, \delta_1)$-differentially private and A_2 and B_2 are $(\varepsilon_2, \delta_2)$-differentially private. Then $A_1 \times A_2$ and $B_1 \times B_2$ are $(\varepsilon_1 + \varepsilon_2, \delta_1 + (1 - \delta_1) \cdot \delta_2)$-differentially private.*

Lemma 4 (Boosting and Differential Privacy (Advanced Composition) [7])**.** *Let $(A_1, B_1), \ldots, (A_k, B_k)$ be pairs of distributions, such that A_i and B_i are (ε, δ)-differentially private for all $i \in \{1, \ldots, k\}$. Then $A_1 \times \ldots \times A_k$ and $B_1 \times \ldots \times B_k$ are $(\hat{\varepsilon}_{\hat{\delta}}, \hat{\delta})$-differentially private, where $\hat{\delta}$ typically is $k \cdot \delta$ and $\hat{\varepsilon}_{\hat{\delta}} = O\left(k\varepsilon^2 + \varepsilon\sqrt{k \log\left(e + (\varepsilon\sqrt{k}/\hat{\delta})\right)}\right)$*

Lemma 5 (Kairouz et al'.s Composition [10])**.** *For any $\varepsilon \geq 0$ and $\delta \in [0, 1]$, the class of (ε, δ)-differentially private mechanisms satisfies*

$$(\varepsilon', \delta')\text{-differential privacy}$$

under k-fold composition, for all $i \in \{0, \ldots, \lfloor k/2 \rfloor\}$ where $\varepsilon' = (k - 2i)\varepsilon$ and $\delta' = 1 - (1 - \delta)^k(1 - \delta_i)$

$$\delta_i = \frac{\sum_{\ell=0}^{i-1} \binom{k}{\ell} \left(e^{(k-\ell)\varepsilon} - e^{(k-2i+\ell)\varepsilon}\right)}{(1 + e^{\varepsilon})^k}$$

These composition results allow for deriving differential-privacy guarantees under composition in a black-box manner, i.e., only depending on ε and δ. Consequently, they are oblivious to how the underlying distributions actually compose and present, in a way, worst-case results under composition. Thus, we cannot expect that they come close to the tight differential privacy guarantee of the composed distributions. In the remainder of this paper we introduce, prove sound and discuss our main idea: approximating the distributions A_1, A_2, B_1, B_2 in a way that allows for a sound calculation of a differential-privacy guarantee that takes into account features of the distribution even under manifold composition. Moreover, we use the same technique to derive a lower bound for the guarantee, to bound the (unknown) tight differential privacy guarantee from both directions.

Buckets for given parameters f and n.

Bucket factor:	f^{-n}	f^{-n+1}	$\cdots$	f^{-2}	f^{-1}	f^0	f^1	f^2	$\cdots$	f^{n-1}	f^n	$> f^n$
Index:	$-n$	$-n+1$	$\cdots$	-2	-1	0	1	2	$\cdots$	$n-1$	n	∞

Figure 1: Depiction of the buckets (separately) constructed for both $\mathcal{B}_A$ and $\mathcal{B}_B$. For $\mathcal{B}_A$ each bucket $\mathcal{B}_A(i)$ with $i \in \{-n+1, \ldots, n\}$ contains all elements $x \in \mathcal{U}$ with $f^{i-1}\Pr[x \leftarrow B] \leq \Pr[x \leftarrow A] \leq f^i \Pr[x \leftarrow B]$, the bucket $\mathcal{B}_A(-n)$ contains all elements with $\Pr[x \leftarrow A] \leq f^{-n}\Pr[x \leftarrow B]$ and the bucket $\mathcal{B}_A(\infty)$ contains all elements with $\Pr[x \leftarrow A] > f^n \Pr[x \leftarrow B]$.

3.4 Bucketing

In this section we introduce the technique of *bucketing a pair of distributions*, by which we mean to derive a precise approximation of the features underlying the pair of distributions that is sufficient for calculating differential privacy and that comes with an efficient way for computing the composition of several such pairs of distributions.

Independence We assume that all distributions A_i, B_i are independent and moreover independent from all distributions A_j, B_j for $i \neq j$. In our composition we acknowledge a certain amount of dependence by composing all distributions A_i with each other and all distributions B_i with each other. Thus, an adversary can indeed gain more information with every step. However, the random choices made by the distributions have to be independent. A result for dependent distributions could be achieved under certain conditions as well, but for the sake of simplicity, we leave such additional complications for future work.

The infinity symbol ∞ In this paper we will write ∞ to describe the corner case accumulated in the largest bucket $\mathcal{B}_\infty$ of our bucket distributions. We consider ∞ to be a distinct symbol and in an abuse of notation, we use the following mathematical rules to interact with it:

- $\infty > i$ for all $i \in \mathbb{Z}$.
- $\infty + i = \infty$ for all $i \in \mathbb{Z}$.

Bucket distribution Our main idea is to divide the universe of all outcomes of the two distributions A and B into sets of elementary events, depending on the ratio between their probabilities in A and in B. If x is an elementary event from the universe $\mathcal{U}$, we consider $\frac{\Pr[x \leftarrow A]}{\Pr[x \leftarrow B]}$ and, depending on this value, decide in which set we put the event. If this fraction is undefined because $\Pr[x \leftarrow B] = 0$, we put the event in a specific set.

The sets we create depend on two parameters: a separation-factor f and a limit n. We then create $2 \cdot n + 2$ sets as follows: For $i \in \{-n, \ldots, n\}$ we assign an elementary event x to the set S_i, if $\Pr[x \leftarrow A] \leq f^i \Pr[x \leftarrow B]$ and if for all $j \in \{-n, \ldots, n\}$ with $j < i$ we have $\Pr[x \leftarrow A] > f^j \Pr[x \leftarrow B]$. All remaining events with $f^n \Pr[x \leftarrow B] < \Pr[x \leftarrow A]$ are assigned to the special set S_∞. For each such set S_i (including for S_∞) we accumulate the probabilities of all events we put into them to yield the respective *bucket* $\mathcal{B}_i$. After creating the buckets we do not need to keep information about the elementary events or the sets anymore: All further calculations are based on our list of $2 \cdot n + 2$ buckets $\mathcal{B}(-n), \ldots, \mathcal{B}(n), \mathcal{B}(\infty)$, which we coin a *bucket distribution*. Thus, the runtime of all further calculations only depends on the number of buckets. After composing two such bucket distributions with each other, we yield another bucketing with the same parameters for f and n. The precision of our method may decrease with the number of compositions, but the complexity of all operations remains the same. We refer to Figure 1 for a graphical presentation of the buckets.

Note that we define a bucket distribution asymmetrically: we consider the probabilities of events occuring in A in relation to the probabilities of the same events occuring in B. We over-approximate the factor between them slightly. Thus, a bucket distribution only delivers guarantees on one direction of differential privacy –

in practice we simply create two bucket distributions: one for relating A with B and one for relating B with A.

Definition 3. *Let A, B be two distributions over the same universe $\mathcal{U}$ and let $f \in \mathbb{R}$ with $f > 1$ and even $n \in \mathbb{N}$ (i.e., there is a $q \in \mathbb{N}$ such that $n = 2q$). Then, $\mathcal{B}(A, B, f, n)$ describes a* bucket distribution $\mathcal{B}$ *over the universe $\{-n, -n+1, \ldots, n\} \cup \{\infty\}$ s.t.*

$$\forall i \in \{-n, -n+1, \ldots, n\} \cup \{\infty\} \,.\mathcal{B}(i) = \sum_{x \in S_i} \Pr\left[x \leftarrow A\right],$$

where the sets S_i are defined as follows:

$$\begin{aligned} S_\infty &= \{x \in \mathcal{U}.\Pr\left[x \leftarrow A\right] > f^n \Pr\left[x \leftarrow B\right]\} \\ \forall i \in \{-n+1, \ldots, n\}\, S_i &= \left\{x \in \mathcal{U}.\ f^{i-1}\Pr\left[x \leftarrow B\right] < \Pr\left[x \leftarrow A\right] \leq f^i \Pr\left[x \leftarrow B\right]\right\} \\ S_{-n} &= \left\{x \in \mathcal{U}.\Pr\left[x \leftarrow A\right] \leq f^{-n}\Pr\left[x \leftarrow B\right]\right\}. \end{aligned}$$

Note that since the sets S_i for $i \in \{-n, \ldots, n\} \cup \{\infty\}$ describe a partitioning of $\mathcal{U}$, we have

$$\sum_{i \in \{-n, \ldots, n\} \cup \{\infty\}} \mathcal{B}(i) = 1.$$

We now define differential privacy for bucket distributions. For all events x in a bucket $\mathcal{B}(i) \neq \mathcal{B}(\infty)$ we know that $\Pr\left[x \leftarrow A\right] \leq f^i \Pr\left[x \leftarrow B\right]$. We over-approximate slightly by treating this inequality as an equality and calculate the delta for differential privacy as in Lemma 1. For the events in $\mathcal{B}(\infty)$ we simply add their full probability to δ, which corresponds to considering these events as total breakdowns of privacy.

Definition 4 (Delta)**.** *Let $f > 1$ and $n \in \mathbb{N}$ and let $\mathcal{B}(A, B, f, n) = \mathcal{B}$ be a bucket distribution. We say that $\mathcal{B}(A, B, f, n)$ is (ε, δ)-DP, if*

$$\sum_{i \in \{-n, \ldots, n\}} \left(\max\left(0, \mathcal{B}(i) \cdot (1 - \frac{e^\varepsilon}{f^i})\right)\right) + \mathcal{B}(\infty) \leq \delta$$

Computing the composition on buckets We proceed by defining how to compose bucket distributions. Our composition shows our main guiding principle behind creating buckets in an exponential manner, described by fractions f^i: Consider the distributions A_1, A_2, B_1, B_2. When composing two buckets $\mathcal{B}_1(i)$ and $\mathcal{B}_2(j)$, we write the result into the bucket $\mathcal{B}_3$ with index $i + j$. The idea behind this strategy is as follows (illustrated in Figure 2). Since events x_1 in $\mathcal{B}_1(i)$ satisfy $\Pr\left[x_1 \leftarrow A_1\right] \leq f^i \Pr\left[x \leftarrow B_1\right]$ and events x_2 in $\mathcal{B}_2(j)$ satisfy $\Pr\left[x_2 \leftarrow A_2\right] \leq f^j \Pr\left[x \leftarrow B_2\right]$, we trivially know that the combined events (x_1, x_2) will satisfy

$$\begin{aligned} \Pr\left[(x_1, x_2) \leftarrow A_1 \times A_2\right] &= \Pr\left[x_1 \leftarrow A_1\right] \cdot \Pr\left[x_2 \leftarrow A_2\right] \\ &\leq f^i \Pr\left[x_1 \leftarrow B_1\right] \cdot f^j \Pr\left[x_2 \leftarrow B_2\right] \\ &= f^{i+j} \Pr\left[(x_1, x_2) \leftarrow B_1 \times B_2\right]. \end{aligned}$$

Following this strategy, we can hence maintain the desired property $\Pr\left[x \leftarrow A\right] \leq f^i \Pr\left[x \leftarrow B\right]$ for all events in bucket i, even after composition. We refer to Figure 3 for a graphical depiction of the bucket composition for one bucket. More formally, we define the composition of two bucket distributions as follows.

Definition 5 (Composition of Buckets)**.** *Let $f > 1$ and $n \in \mathbb{N}$ and let $\mathcal{B}(A_1, B_1, f, n) = \mathcal{B}_1$ and $\mathcal{B}(A_2, B_2, f, n) = \mathcal{B}_2$ be two bucket distributions. We define the composition of the pairs as $\mathcal{B}_1 \times \mathcal{B}_2$, where*

$$\forall i \in \{-n, -n+1, \ldots, n\} \cup \{\infty\} \,.\mathcal{B}_1 \times \mathcal{B}_2(i) = \sum_{(j,k) \in S_i} \mathcal{B}_1(j) \cdot \mathcal{B}_2(k),$$

where the sets S_i are defined as follows:

$$\begin{aligned} S_\infty &= \left\{(j, k) \in \{-n, \ldots, n, \infty\}^2 \,.j + k > n\right\} \\ \forall i \in \{-n+1, \ldots, n\}\, S_i &= \left\{(j, k) \in \{-n, \ldots, n\}^2 \,.j + k = i\right\} \\ S_{-n} &= \left\{(j, k) \in \{-n, \ldots, n\}^2 \,.j + k \leq -n\right\}. \end{aligned}$$

Bucket composition example for two events for $n = 5$.

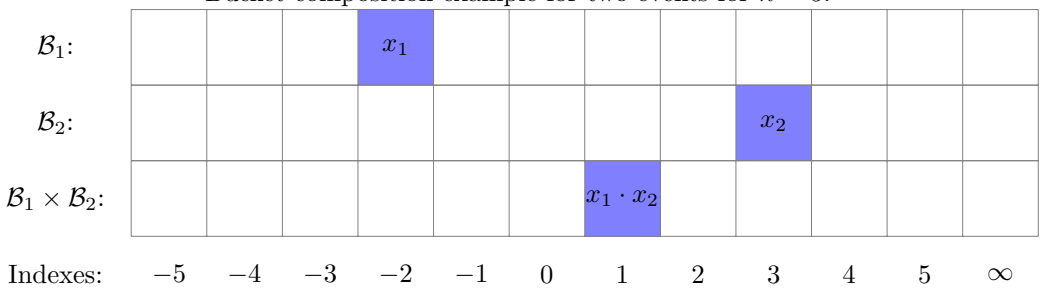


Figure 2: Depiction of how individual events x_1 and x_2 compose into new buckets.

Bucket composition example for bucket indes 1 for $n = 5$, only showing the bucket indexes.

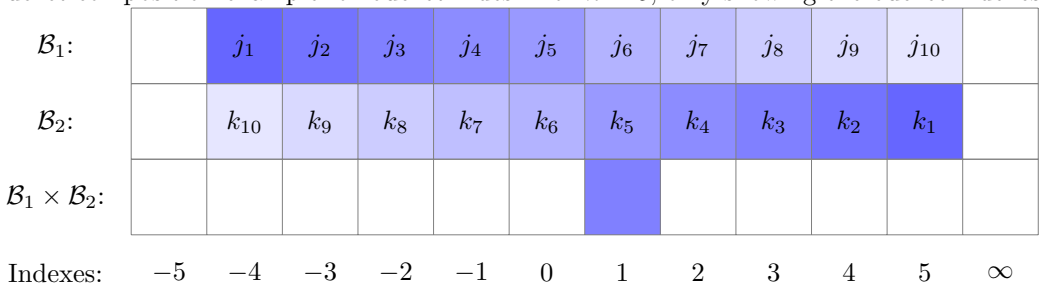


Figure 3: Depiction of the bucket composition for the (new) bucket with index $i = 1$. We calculate the value of the bucket i by summing over the product of all $\mathcal{B}_1(j_i) \cdot \mathcal{B}_2(k_i)$. Graphically, buckets with the same color are combined. Note that none of the buckets ∞ and -5 are used for the composition, as for all $j \in \{-5, \ldots, 5\}$, $\infty + j \neq 1$ and $-5 + j \neq 1$.

Note that since the sets S_i for $i \in \{-n, \ldots, n\} \cup \{\infty\}$ describe a partitioning of $\mathcal{U}$ and since additionally the buckets $\mathcal{B}_1$ and $\mathcal{B}_2$ add up to 1, we have

$$\sum_{i \in \{-n, \ldots, n\} \cup \{\infty\}} \mathcal{B}_1 \times \mathcal{B}_2(i) = 1.$$

When composing bucket distributions, they naturally "broaden", i.e., more and more events populate buckets that are further away from the middle bucket with factor f^0. When creating a bucket distribution for a given number n, this effect leads to a trade-off between the granularity (i.e., the choice of the bucket factor f) and the expected number of compositions: the smaller f, the more precise does the bucket distribution model the features of the distributions, but the fewer compositions before a significant amount of events reaches the corner buckets ($\mathcal{B}(-n)$ and $\mathcal{B}(\infty)$), which again reduces the precision. To counter this (rather annoying) effect, we introduce an additional operation which we coin *squaring*: we square the factor f, thus halving the precision of the buckets, and soundly merge buckets into this new, more coarse-grained bucket distribution. Squaring allows us to start with a much more fine-grained bucket distribution and reduce the granularity as we compose, which can improve the overall precision of the approach significantly. We choose to square f instead of increasing it to an arbitrary f' to ease the computation of the new bucket distribution: we simply combine buckets $2i - 1$ and $2i$ with factors f^{2i-1} and f^{2i} into the new bucket i with factor $(f^2)^i = f^{2i}$. We refer to Figure 4 for a graphical depiction of squaring.

The composition of bucket distributions is commutative, but not associative and the number of times and also the times at which the squaring was performed are relevant as well. Hence, we need to keep track of the order in which we applied composition and squaring. To this end, we define *composition trees* that are important for our proofs, but not for calculating actual results (since we show that any composition tree leads to sound results), and can thus be considered a purely technical definition.

Definition 6 (Composition trees)**.** *For two sets of tuples $(A_1, \ldots, A_\mathcal{W})$ and $(B_1, \ldots, B_\mathcal{W})$ of the same size u, a* composition tree over $(A_1, \ldots, A_\mathcal{W})$ and $(B_1, \ldots, B_\mathcal{W})$ *is a tree with three kinds of nodes that are all*

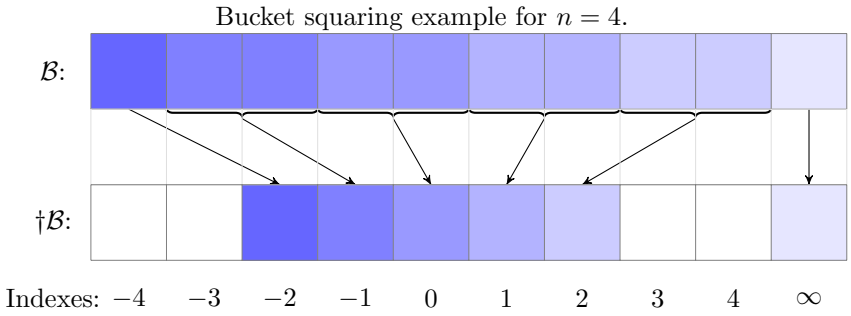


Figure 4: Depiction of the bucket squaring. Events from each bucket $\mathcal{B}(i)$ are moved into bucket $\mathcal{B}(\lceil i/2 \rceil)$, with the exception of $\mathcal{B}(\infty)$, which remains unchanged.

labeled with a bucket factor $f > 1$; leaves ($T = \mathbb{P}(A_i, B_i)$) are additionally labelled with a pair of distributions, composition nodes ($T = T_1 \times T_2$) with exactly two child nodes and squaring nodes ($T = \dagger T_1$) with exactly one child node. We require that each pair of distributions (A_i, B_i) is the label of exactly one leaf, that for each composition node the child nodes have the same f in the label, and that the label of each squaring node contains f^2 if the child node's label contains f.

For ease of notation we write (A_i, B_i) to describe the tree consisting only of a leaf $\mathbb{P}(A_i, B_i)$. For brevity, we even write A_i or B_i for the same tree, if we only talk about the respective distribution.

For discussing our results and the soundness of our results we want to compare the differential privacy guarantees of bucket distributions with the real differential privacy guarantees (calculating which might not be feasible). To this end and for talking about individual elementary events we assign an index to each such event. The index specifies the (one) bucket the respective event influences. For a bucket distribution that has been created from distributions (and not composed), this index is simply the bucket the event was assigned to. After composition, the index depends on how the indexes of the respective buckets interacted: in the most simple case, if x_1 and x_2 are events with indexes i and j, then the event (x_1, x_2) will have the index $i + j$. However, the corner cases can modify the index, as the index can only be in the set $\{-n, \ldots, n, \infty\}$.

Definition 7 (Index of an event according to buckets)**.** *Let $(A_k, B_k)_{k=1}^{\mathcal{W}}$ be pairs of distributions over the universes $(\mathcal{U}_i)_{k=1}^{\mathcal{W}}$, let $f > 1$ and $n \in \mathbb{N}$.*

We define the set of indexes for events $x = (x_1, \ldots, x_{\mathcal{W}}) \in (\mathcal{U}_i)_{k=1}^{\mathcal{W}}$ as follows. First, we define for the individual components $x_k \in \mathcal{U}_k$ with $k \in \{1, \ldots, \mathcal{W}\}$,

$$i_{\mathbb{P}(A_k,B_k)}(x_k) := \begin{cases} l & \text{if } l \in \{-n+1, \ldots, n\} \\ & \wedge f^{l-1}\Pr[x_k \leftarrow B_k] < \Pr[x_k \leftarrow A_k] \leq f^l \Pr[x_k \leftarrow B_k] \\ \infty & \text{if } \Pr[x_k \leftarrow A_k] > f^n \Pr[x_k \leftarrow B_k] \\ -n & \text{otherwise} \end{cases}$$

For any pair of composition trees T_1, T_2 over some probability distributions, and

- *for $T = T_1 \times T_2$ we define the index of $x = (x_1, x_2)$ as*

$$i_T(x) = i_{T_1 \times T_2}(x_1, x_2) := \begin{cases} -n & \text{if } i_{T_1}(x_1) + i_{T_2}(x_2) < -n \\ \infty & \text{if } i_{T_1}(x_1) + i_{T_2}(x_2) > n \\ i_{T_1}(x_1) + i_{T_2}(x_2) & \text{otherwise,} \end{cases}$$

  where we assume that $\forall y, z \in \mathbb{Z},\ y + \infty = \infty > z$.

- *for $T = \dagger T_1$ we define the index of x as*

$$i_T(x) = i_{\dagger T_1}(x) := \begin{cases} \lceil i_{T_1}(x)/2 \rceil & \text{if } i_{T_1}(x) \neq \infty \\ \infty & \text{otherwise,} \end{cases}$$

We stress that $i_{T_1 \times T_2}(x_1, x_2)$ is not necessarily associative, i.e., there are distributions $A_1, A_2, A_3, B_1, B_2, B_3$, and x_1, x_2, x_3 such that

$$i_{(T_1 \times T_2) \times T_3}(x_1, x_2, x_3) \neq i_{T_1 \times (T_2 \times T_3)}(x_1, x_2, x_3)$$

Soundness of differential privacy guarantees for bucket distributions We can now start to argue about the differential privacy guarantees we calculate for bucket distributions. We will show that if a bucket distribution is (ε, δ)-differentially private, then the distributions from which the pair was created (either directly or via composition) is also (ε, δ)-differentially private. Simply put, the guarantees we calculate are sound.

We begin by showing a helpful lemma that directly follows our main strategy: all elementary events x that are assigned an index $i \neq \infty$ (according to a composition tree T) satisfy $\Pr[x \leftarrow A] \leq f^i \Pr[x \leftarrow B]$.

Lemma 6. *Let $(A_k, B_k)_{k=1}^{\mathcal{W}}$ be pairs of distributions over the universes $(\mathcal{U}_i)_{k=1}^{\mathcal{W}}$, let $n \in \mathbb{N}$ and let for all $k \in \{1, \ldots, u\}$. Let $A := \prod_{k=1}^{\mathcal{W}} A_k$ and $B := \prod_{k=1}^{\mathcal{W}} B_k$.*

For all $x \in \prod_{k=1}^{\mathcal{W}} \mathcal{U}_k$ and for every composition tree T over $A_1, \ldots, A_{\mathcal{W}}$ such that $i_T(x) \neq \infty$ and the root node has f in the label, we have $\Pr[x \leftarrow A] \leq f^{i_T(x)} \Pr[x \leftarrow B]$. Analogously, with $i_{T^B}(x) \neq \infty$ we have $\Pr[x \leftarrow B] \leq f^{i_T(x)} \Pr[x \leftarrow A]$.

Proof. We show the lemma by a structural induction over the composition tree.

- For the leaves (i.e., $T_k = \text{⨀}(A_k, B_k)$), if $i_T(x) \neq \infty$, then for all $k \in \{1, \ldots, k\}$ $i_{\text{⨀}(A_k, B_k)}(x_k) \neq \infty$.
  For each k such that $i_{\text{⨀}(A_k, B_k)}(x_k) = -n$, from the case distinction of $i_{\text{⨀}(A_k, B_k)}(x_k)$ in Definition 7 it follows that

  $$\Pr[x_k \leftarrow A_k] \leq f^{-n} \Pr[x_k \leftarrow B_k] \tag{1}$$
  $$\Pr[x_k \leftarrow A_k] \leq f^{i_{\text{⨀}(A_k, B_k)}(x_k)} \Pr[x_k \leftarrow B_k]. \tag{2}$$

  Hence, we get from Definition 7 and Equation (2) that for all k such that $i_{\text{⨀}(A_k, B_k)}(x_k) \neq \infty$, we have

  $$\Pr[x_k \leftarrow A_k] \leq f^{i_{\text{⨀}(A_k, B_k)}(x_k)} \Pr[x_k \leftarrow B_k]. \tag{3}$$

- For composition nodes (i.e., $T = T_1 \times T_2$), where both children are labelled with f (and consequently the composition node is also labelled with f), we get for x

  $$\begin{aligned}
  \Pr[x \leftarrow A] &= \prod_{k=1}^{\mathcal{W}} \underbrace{\Pr[x_k \leftarrow A_k]}_{\leq f^{i_{\text{⨀}(A_k, B_k)}(x_k)} \Pr[x_k \leftarrow B_k]} \\
  &\leq \prod_{k=1}^{\mathcal{W}} f^{i_{\text{⨀}(A_k, B_k)}(x_k)} \Pr[x_k \leftarrow B_k] = \underbrace{\prod_{k=1}^{\mathcal{W}} f^{i_{\text{⨀}(A_k, B_k)}(x_k)}}_{\leq f^{i_T(x)}} \underbrace{\prod_{k=1}^{\mathcal{W}} \Pr[x_k \leftarrow B_k]}_{= \Pr[x \leftarrow B]} \\
  &\leq f^{i_T(x)} \Pr[x \leftarrow B]
  \end{aligned}$$

  Note that $\sum_{k \in \{1, \ldots, \mathcal{W}\}} i_{\text{⨀}(A_k, B_k)} \leq i_T(x)$ holds by definition of the index over any composition tree: at every node at least the sum of the underlying nodes is considered (or $-n$ if that sum is $< -n$).

- For squaring nodes (i.e., $T = \dagger T_1$), where the child node is labelled with f (and the re-scale node thus is labelled with f^2), we know that $i_T(x) \neq \infty \equiv i_{T_1}(x) \neq \infty$. For $i_{T_1}(x) \neq \infty$ we know by the induction hypothesis that $\Pr[x \leftarrow A] \leq f^{i_{T_1}(x)} \Pr[x \leftarrow B]$. By definition, we have

  $$\begin{aligned}
  \Pr[x \leftarrow A] &\leq f^{i_{T_1}(x)} \Pr[x \leftarrow B] \\
  &= f^{2 i_{T_1}(x)/2} \Pr[x \leftarrow B] \\
  &\leq (f^2)^{\lceil i_{T_1}(x)/2 \rceil} \Pr[x \leftarrow B] \\
  &= (f^2)^{i_T(x)} \Pr[x \leftarrow B]
  \end{aligned}$$

□

We now state the first theorem of our paper: the buckets are sound.

Theorem 1 (Buckets are sound)**.** *Let A, B be two distributions over the same universe $\mathcal{U}$, let $f > 1$ and $n \in \mathbb{N}$ and let $\mathcal{B}(A, B, f, n) = \mathcal{B}$ be a bucket distribution. If for $\varepsilon, \delta \geq 0$, $\mathcal{B}(A, B, f, n)$ is (ε, δ)-DP, then A and B are (ε, δ)-differentially private.*

Proof. The theorem can be shown directly; however, as it follows quite trivially from the proof of a more complicated case we consider in the subsequent chapter, we omit the proof here and refer to Lemma 12 instead. □

Corollary 1. *For any privacy-enhancing mechanism M for which there exist worst-case inputs x_0, x_1, let $\mathcal{B}(M(x_0), M(x_1), f, n) = \mathcal{B}$ be a bucket distribution. If for $\varepsilon, \delta \geq 0$, $\mathcal{B}(M(x_0), M(x_1), f, n)$ is (ε, δ)-DP, then M is (ε, δ)-differentially private. Moreover, if $\mathcal{B}(M(x_0), M(x_1), f, n)^k$ is (ε, δ)-DP then then M is (ε, δ)-differentially private under k-fold composition.*

As described in Section 2, distributions can be used to calculate differential privacy in a variety of applications. Technically, we require the existence of *worst-case* inputs that allow us to directly derive the relevant distributions.

Definition 8 (Worst-case inputs)**.** *Inputs x_0, x_1 are worst-case inputs for a given sensitivity s and a mechanism M if $\Pr[M(x_0) \in S] \leq e^{\varepsilon} \Pr[M(x_1) \in S] + \delta$, then M is (ε, δ)-DP for all inputs with sensitivity s. Such worst-case inputs typically exist for practical privacy-preserving mechanisms that achieve (ε, δ)-DP.*

Proof. Consider the reduction that replaces all inputs of the attacker with sensitivity s with the worst case inputs for sensitivity s. If there were inputs x_0', x_1' such that for any ε

$$\Pr[M(x_0') \in S] \geq e^{\varepsilon} \Pr[M(x_1') \in S] + \delta',$$

although

$$\Pr[M(x_0) \in S] \leq e^{\varepsilon} \Pr[M(x_1) \in S] + \delta$$

and $\delta' > \delta$, then x_0, x_1 cannot be the worst-case inputs. □

Our approach can be applied whenever worst-case inputs for instance of the mechanism can be found independently of the random coins used by the mechanism in the previous rounds. This is commonly the case when differential privacy is applied.

4 Reducing and bounding the error

We have already presented a sound way of approximating a distribuion pair by creating a bucket distribution. Our calculations from the previous section lead to sound and, in many cases, better results than generic composition theorems from the literature. In this section we explore the precision of our results: we define error (correction) terms that help us to both find a lower bound on the differential privacy guarantee for the considered distributions even under manifold composition, and to find a tighter guarantee for differential privacy.

We distinguish between two types of error correction terms: the *real error correction term* ℓ that captures the value we use to tighten our result in a sound way and the *virtual error correction term* $\tilde{\ell}$ that captures the maximal influence an error correction term can have. The virtual error correction term accurately captures the difference between the probability an event x appears to have in the alternative distribution (using the bucket factor) $\frac{\Pr[x \leftarrow A]}{f^i}$ and the probability that it actually has in the alternative $\Pr[x \leftarrow B]$. In some cases, however, we misplace an event, s.t., it ends up an a bucket with an index that is too large: events x that should not be considered for the overall guarantee, i.e., that have $\Pr[x \leftarrow A] - e^{\varepsilon} \Pr[x) \leftarrow B] < 0$ can appear in a bucket with index i s.t. $e^{\varepsilon} < f^i$. Thus, correctly calculating the error correction term while possibly misplacing events can lead to wrong results.

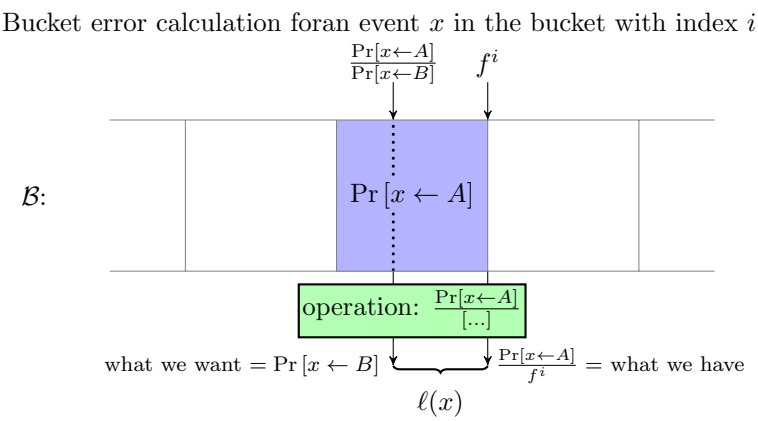


Figure 5: Depiction of how and why we calculate the error. We only preserve f^i and (accumulated) $\Pr[x \leftarrow A]$, but would like to preserve $\Pr[x \leftarrow B]$, which we can approximately get by: $\frac{\Pr[x \leftarrow A]}{f^i}$ and thus we store the difference between those values in the error term.

There are two reasons for why events can be misplaced: First, when composing bucket distributions, events can be misplaced by one bucket. We take care of this by not including the error correction terms of a certain number of buckets, depending on the number of compositions. Second, when events are put into the smallest bucket (with index $-n$), they can be arbitrarily "misplaced", particularly after a composition. To counter this effect, we introduce the real error correction term, in which we do not include the error of the smallest bucket (with index $-n$).

4.1 Buckets with error correction terms

Our strategy is as follows, assuming two distributions A and B: Whenever we enter an event x into a bucket $\mathcal{B}(i)$, we remember the difference between the probability that the event occurs in A, adjusted by the bucket factor, and the probability that the same event occurs in B: $\ell(i) += \Pr[x \leftarrow B] - \frac{\Pr[x \leftarrow A]}{f^i}$. Recall that the main purpose of the buckets is to keep track of the ratio between those two probabilities. We sum up all these *error correction terms* per individual bucket.

Let us, for the sake of illustration, consider one bucket $\mathcal{B}(i)$, containing events $x \in S_i$ for a set S_i.

$$\frac{\mathcal{B}(i)}{f^i} - \ell(i) = \frac{\sum_{x \in S_i} \Pr[x \leftarrow A]}{f^i} - \sum_{x \in S_i} \left(\Pr[x \leftarrow B] - \frac{\Pr[x \leftarrow A]}{f^i} \right) = \sum_{x \in S_i} \Pr[x \leftarrow B].$$

Thus, only considering one additional value per bucket, we can precisely remember the probability that the events occurred in B and we can then use this probability to calculate a more precise differential privacy guarantee. We ommit the error correction terms for the bucket $\mathcal{B}(\infty$, as there is no bucket factor attached to it (so there is no value the error correction term could correct).

We later see that given a value for ε we need to be careful when dealing with exactly one bucket: the bucket $\mathcal{B}(j)$ with $f^{j-1} < e^{\varepsilon} \leq f^j$. If we were precise in our calculations, we would only consider *some* of the events from the bucket, namely the ones with $\Pr[x \leftarrow A] \leq e^{\varepsilon} \Pr[x \leftarrow B]$, but since we combined them all into one bucket, we cannot distinguish the individual events anymore. To retain a sound guarantee, we don't consider the error correction term of this bucket when calculating δ. Under composition this error slightly increases, as events can be "misplaced" by more than one bucket when we compose the buckets. Consequently, every composition increases the number of buckets for which we don't consider an error correction term.

Definition 9 (Bucket distribution with error correction terms)**.** *Let A, B be a pair of distributions over the universes $\mathcal{U}$, let $f > 1$ and $n \in \mathbb{N}$ and let for all $k \in \{1, \ldots, \mathcal{W}\}$*

We define the bucket distribution with error correction terms $\mathcal{B}(A, B, f, n) = (\mathcal{B}, \tilde{\ell}, \ell, f, 1)$*, as*

$$\begin{aligned}&\forall i \in \{-n, \ldots, n, \infty\}\\ \mathcal{B}(i) :=& \sum_{x \in \mathcal{U} \ s.t. \ i(x)=i} \Pr\left[x \leftarrow A\right]\end{aligned}$$

where we define the error correction terms as

$$\begin{aligned}&\forall i \in \{-n, \ldots, n\}\\ \tilde{\ell}(i) :=& \sum_{x \in \mathcal{U} \ s.t. \ i(x)=i} \Pr\left[x \leftarrow B\right] - \frac{\Pr\left[x \leftarrow A\right]}{f^i}\\ &\tilde{\ell}(\infty) := 0\\ &\forall i \in \{-n+1, \ldots, n\}\\ \ell(i) :=& \tilde{\ell}(i)\\ &\ell(-n) := \ell(\infty) := 0.\end{aligned}$$

For completeness we re-define the composition and squaring of buckets first (which is unchanged from the previous section) and then define how the error terms behave under both composition and squaring: under composition, we want to calculate a perfect error correction term for the combined events, i.e., given events x_1 and x_2 with (individual) error terms $\Pr\left[x_1 \leftarrow B_1\right] - \frac{\Pr[x \leftarrow A_1]}{f^{i_1}}$ and $\Pr\left[x_2 \leftarrow B_2\right] - \frac{\Pr[x_2 \leftarrow A_2]}{f^{i_2}}$ we want (in the typical case, ignoring corner cases) to have an error correction term for the pair of the form $\Pr\left[(x_1, x_2) \leftarrow B_1 \times B_2\right] - \frac{\Pr[(x_1, x_2) \leftarrow A_1 \times A_2]}{f^{i_1+i_2}}$. However, the buckets cannot keep track of the value for $\Pr\left[(x_1, x_2) \leftarrow B_1 \times B_2\right]$ – recall that this is precisely why we have introduced the error terms. Fortunately, we can calculate the desired error correction terms from the previous error correction terms, the bucket values, and the bucket factors.

Similarly, for the squaring, we quantify how the error terms change when we modify the buckets. Although each new bucket is composed of two previous buckets, the bucket factor actually only changes for one half of the values: the evenly indexed buckets $\mathcal{B}(2i)$ with factor f^{2i} are now moved into buckets $\mathcal{B}(i)$ with the same factor $(f^2)^i$ and thus there error correction terms are still correct. The other half of buckets $\mathcal{B}(2i-1)$ with factor f^{2i-1} are moved into the same buckets $\mathcal{B}(i)$ with factor $(f^2)^i$ and thus the error correction terms need to be modified to capture this change in the bucket factor.

We define the composition and squaring as follows.

Definition 10 (Composition and squaring with error correction terms)**.** *For two bucket distributions* $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ *over a universe* $\prod_{k=1}^{\mathcal{W}_1} \mathcal{U}_k$ *and* $(\mathcal{B}_2, \tilde{\ell}_2, \ell_2, f_2, u_2)$ *over a universe* $\prod_{k=\mathcal{W}_1+1}^{\mathcal{W}_1+\mathcal{W}_2} \mathcal{U}_k$*, with* $f_1 = f_2 = f$*, we have*

$$\begin{aligned}&(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1) \times (\mathcal{B}_2, \tilde{\ell}_2, \ell_2, f_2, u_2)\\ :=&(\mathcal{B}_1 \times \mathcal{B}_2, \tilde{\ell}_1 \times \tilde{\ell}_2, \ell_1 \times \ell_2, f, u_1 + u_2)\end{aligned}$$

where $\mathcal{B}_X \times \mathcal{B}_Y$ *(for* $X \in \{A_1, A_2\}$ *and* $Y \in \{B_1, B_2\}$*) is defined as*

$$\begin{aligned}\mathcal{B}_1 \times \mathcal{B}_2(-n) :=& \sum_{j,k \in \{-n,\ldots,n\}^2 \ s.t. \ j+k \leq -n} \mathcal{B}_1(j) \cdot \mathcal{B}_2(k)\\ \forall i \in \{-n+1, \ldots, n\} \ \ \mathcal{B}_1 \times \mathcal{B}_2(i) :=& \sum_{j,k \in \{-n,\ldots,n\}^2 \ s.t. \ j+k=i} \mathcal{B}_1(j) \cdot \mathcal{B}_2(k)\\ \mathcal{B}_1 \times \mathcal{B}_2(\infty) :=& \sum_{j,k \in \{-n,\ldots,n,\infty\}^2 \ s.t. \ j+k > -n} \mathcal{B}_1(j) \cdot \mathcal{B}_2(k)\end{aligned}$$

where we define the error correction terms as

$$\begin{aligned}
&\forall i \in \{-n+1, \ldots, n\} \\
\tilde{\ell}_1 \times \tilde{\ell}_2(i) &:= \sum_{(k,l)\in\{-n,\ldots,n\}^2, k+l=i} \left(\frac{\mathcal{B}_1(k)}{f^k} + \tilde{\ell}_1(k)\right)\tilde{\ell}_2(l) + \tilde{\ell}_1(k)\left(\frac{\mathcal{B}_2(l)}{f^l} + \tilde{\ell}_2(l)\right) - \tilde{\ell}_1(k)\tilde{\ell}_2(l) \\
\tilde{\ell}_1 \times \tilde{\ell}_2(-n) &:= \sum_{(k,l)\in\{-n,\ldots,n\}^2, k+l\leq -n} \left(\frac{\mathcal{B}_1(k)}{f^k} + \tilde{\ell}_1(k)\right)\tilde{\ell}_2(l) + \tilde{\ell}_1(k)\left(\frac{\mathcal{B}_2(l)}{f^l} + \tilde{\ell}_2(l)\right) - \tilde{\ell}_1(k)\tilde{\ell}_2(l) \\
&\tilde{\ell}_1 \times \tilde{\ell}_2(\infty) := 0 \\
&\forall i \in \{-n, \ldots, n\} \\
\ell_1 \times \ell_2(i) &:= \sum_{(k,l)\in\{-n,\ldots,n\}^2, k+l=i} \left(\frac{\mathcal{B}_1(k)}{f^k} + \ell_1(k)\right)\ell_2(l) + \ell_1(k)\left(\frac{\mathcal{B}_2(l)}{f^l} + \ell_2(l)\right) - \ell_1(k)\ell_2(l) \\
&\ell_1 \times \ell_2(-n) := \ell_1 \times \ell_2(\infty) := 0.
\end{aligned}$$

For a bucket distributions $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ *over a universe* $\prod_{k=1}^{\mathcal{W}_1} \mathcal{U}_k$*, we have*

$$\begin{aligned}
&\dagger(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1) \\
:=&(\dagger\mathcal{B}_1, \dagger\tilde{\ell}_1, \dagger\ell_1, f_1^2, \lceil u_1/2 \rceil + 1)
\end{aligned}$$

where we define

$$\begin{aligned}
\dagger\mathcal{B}_1(-n/2) &:= \mathcal{B}_1(-n) \\
\forall i \in \{-n/2+1, \ldots, n/2\}\ \dagger\mathcal{B}_1(i) &:= \mathcal{B}_1(2\cdot i - 1) + \mathcal{B}_1(2\cdot i) \\
\dagger\mathcal{B}_1(\infty) &:= \mathcal{B}_1(\infty) \\
\forall i \in \{-n, \ldots, -n/2-1, n/2+1, \ldots, n\}\ \dagger\mathcal{B}_1(i) &:= 0
\end{aligned}$$

where we define the error correction terms as

$$\begin{aligned}
&\forall i \in \{-n/2+1, \ldots, n/2\} \\
\dagger\tilde{\ell}_1(i) &:= \tilde{\ell}_1(2i-1) + \mathcal{B}_1(2i-1)\left(\frac{1}{f_1^{2i-1}} - \frac{1}{f_1^{2i}}\right) + \tilde{\ell}_1(2i) \\
\dagger\tilde{\ell}_1(-n/2) &:= \tilde{\ell}_1(-n) \\
\forall i \in \{-n, \ldots, -n/2-1, n/2+1, \ldots, n, \infty\}\ \dagger\tilde{\ell}_1(i) &:= 0 \\
&\forall i \in \{-n/2+1, \ldots, n/2\} \\
\dagger\ell_1(i) &:= \ell_1(2i-1) + \mathcal{B}_1(2i-1)\left(\frac{1}{f_1^{2i-1}} - \frac{1}{f_1^{2i}}\right) + \ell_1(2i) \\
\forall i \in \{-n, \ldots, -n/2, n/2+1, \ldots, n, \infty\}\ \dagger\ell_1(i) &:= 0.
\end{aligned}$$

To improve the readability of our proofs we introduce a more compact notation for bucket distributions that stem from a composition tree, by slightly abusing the $\prod$ symbol.

Definition 11 (Notation for composing bucket distributions)**.** *Given a composition tree* $T = \text{⊘}(A, B)$ *over the distributions* A *and* B*, we write*

$$\prod_{k\in\{1\}}^{T} \mathcal{B}(A_k, B_k, f_k, n) = \mathcal{B}(A, B, f, n).$$

Given a composition tree $T = T_1 \times T_2$, where T_1 is over the distributions $(A_1, \dots, A_j)$ and $(B_1, \dots, B_j)$ and T_2 is over the distributions $(A_{j+1}, \dots, A_{\mathcal{W}})$ and $(B_{j+1}, \dots, B_{\mathcal{W}})$, we write

$$\prod_{k\in\{1,\dots,\mathcal{W}\}}^{T} \mathcal{B}(A_k, B_k, f_k, n) = \left(\prod_{k\in\{1,\dots,j\}}^{T_1} \mathcal{B}(A_k, B_k, f_k, n)\right) \times \left(\prod_{k\in\{j+1,\dots,\mathcal{W}\}}^{T_2} \mathcal{B}(A_k, B_k, f_k, n)\right).$$

Given a composition tree $T = \dagger T_1$, where T_1 is over the distributions $(A_1, \dots, A_{\mathcal{W}})$, we write

$$\prod_{k\in\{1,\dots,\mathcal{W}\}}^{T} \mathcal{B}(A_k, B_k, f_k, n) = \dagger \left(\prod_{k\in\{1,\dots,j\}}^{T_1} \mathcal{B}(A_k, B_k, f_k, n)\right).$$

Whenever we say that a bucketing $(\mathcal{B}, \tilde{\ell}, \ell, f, u)$ over a universe $\prod_{k=1}^{\mathcal{W}} \mathcal{U}_k$ is defined for a value n and with a composition tree T, we mean

$$(\mathcal{B}, \tilde{\ell}, \ell, f, u) := \prod_{k\in\{1,\dots,\mathcal{W}\}}^{T} \mathcal{B}(A_k, B_k, f_k, n),$$

where $(\mathbb{P}(A_1, B_1), \dots, \mathbb{P}(A_{\mathcal{W}}, B_{\mathcal{W}}))$ are the leaf nodes of T.

4.2 Buckets and error correction terms per element

Before we can show the first helpful lemmas for the soundness of our error correction terms, we introduce the impact that each individual event x has on the bucket terms that are influenced by x. We first simply define these terms per element separately and then continue by showing that each bucket value (and error correction term) is simply the sum over the respective terms of all elements contributing to this bucket. This marks a significant step in the correctness (and tightness) of our results: Although we only consider a few values (one bucket value and one error correction value per bucket) we still capture all individual events. The only exception to this precision then comes from misplaced events, which we will analyze subsequently.

Definition 12 (Bucket distribution with error correction terms per element)**.** *Let A, B be a pair of distributions over the universes $\mathcal{U}$, let $f > 1$ and $n \in \mathbb{N}$ and let for all $k \in \{1, \dots, \mathcal{W}\}$*

We define the bucket distribution with error correction terms $\mathcal{B}(A, B, f, n) = (\mathcal{B}, \tilde{\ell}, \ell, f, 1)$, as follows

$$\begin{aligned}
\mathcal{B}(x) &:= \Pr[x \leftarrow A] \\
\text{if } i_{\mathbb{P}(A,B)}(x) \in \{-n, \dots, n\},\ \tilde{\ell}(x) &:= \Pr[x \leftarrow B] - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} \\
\text{if } i_{\mathbb{P}(A,B)}(x) = \infty,\ \tilde{\ell}(x) &:= 0 \\
\text{if } i_{\mathbb{P}(A,B)}(x) \in \{-n+1, \dots, n, \infty\},\ \ell(x) &:= \tilde{\ell}(x) \\
\text{if } i_{\mathbb{P}(A,B)}(x) = -n,\ \ell(x) &:= 0.
\end{aligned}$$

Both the composition and squaring for our terms per element behave identically to the corresponding terms per bucket. The only difference here is that we rely on the index per element i_T instead of the bucket indexes.

Definition 13 (Composition with error correction terms per element)**.** *For two bucket distributions $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ over a universe $\prod_{k=1}^{\mathcal{W}_1} \mathcal{U}_k$ and $(\mathcal{B}_2, \tilde{\ell}_2, \ell_2, f_2, u_2)$ over a universe $\prod_{k=\mathcal{W}_1+1}^{\mathcal{W}_1+\mathcal{W}_2} \mathcal{U}_k$, both defined with the same values f and n, and with composition trees T_1 and T_2 we have for each $x = (x_1, x_2) \in \prod_{k=1}^{\mathcal{W}_1} \mathcal{U}_\times \prod_{k=\mathcal{W}_1+1}^{\mathcal{W}_1+\mathcal{W}_2} \mathcal{U}_k$,*

$$\mathcal{B}_1 \times \mathcal{B}_2(x) := \mathcal{B}_1(x_1) \cdot \mathcal{B}_2(x_2)$$

and we define the error correction terms as

$$\begin{aligned}
&\text{if } i_{T_1\times T_2}(x) \in \{-n,\dots,n\}\\
\tilde{\ell}_1 \times \tilde{\ell}_2(x) &:= \left(\frac{\mathcal{B}_1(x_1)}{f^{i_{T_1}(x_1)}} + \tilde{\ell}_1(x_1)\right)\tilde{\ell}_2(x_2) + \tilde{\ell}_1(x_1)\left(\frac{\mathcal{B}_2(x_2)}{f^{i_{T_2}(x_2)}} + \tilde{\ell}_2(x_2)\right) - \tilde{\ell}_1(x_1)\tilde{\ell}_2(x_2)\\
&\text{if } i_{T_1\times T_2}(x) \in \{\infty\}\\
&\tilde{\ell}_1 \times \tilde{\ell}_2(x) := 0\\
&\text{if } i_{T_1\times T_2}(x) \in \{-n+1,\dots,n,\infty\}\\
\ell_1 \times \ell_2(x) &:= \left(\frac{\mathcal{B}_1(x_1)}{f^{i_{T_1}(x_1)}} + \ell_1(x_1)\right)\ell_2(x_2) + \ell_1(x_1)\left(\frac{\mathcal{B}_2(x_2)}{f^{i_{T_2}(x_2)}} + \ell_2(x_2)\right) - \ell_1(x_1)\ell_2(x_2)\\
&\text{if } i_{T_1\times T_2}(x) \in \{-n,\infty\}\\
&\ell_1 \times \ell_2(x) := 0.
\end{aligned}$$

For a squaring node ($T = \dagger T_1$), we keep the bucket value as $\dagger\mathcal{B}_1(x) := \mathcal{B}_1(x_1)$ and we define the error correction terms as follows (where f is the old factor, from the label of T_1):

$$\begin{aligned}
&\text{if } i_{T_1}(x) \in \{-n,\dots,n\}\\
\dagger\tilde{\ell}_1(x) &:= \tilde{\ell}_1(x) + \mathcal{B}_1(x)\cdot\left(\frac{1}{f^{i_{T_1}(x)}} - \frac{1}{f^{2\cdot\lceil i_{T_1}(x)/2\rceil}}\right)\\
&\text{if } i_{T_1}(x) \in \{\infty\}\\
&\dagger\tilde{\ell}_1(x) := 0\\
\\
&\text{if } i_{T_1}(x) \in \{-n+1,\dots,n\}\\
\dagger\ell_1(x) &:= \ell_1(x) + \mathcal{B}_1(x)\cdot\left(\frac{1}{f^{i_{T_1}(x)}} - \frac{1}{f^{2\cdot\lceil i_{T_1}(x)/2\rceil}}\right)\\
&\text{if } i_{T_1}(x) \in \{-n,\infty\}\\
&\dagger\ell_1(x) := 0.
\end{aligned}$$

We now show our first important lemma for the soundness of our buckets and error correction terms: the terms we defined just previously indeed characterize the impact of each individual event on the overall bucket values and error correction terms. These terms indeed are just the sum of the respective values per element for all elements of an index that equals the bucket index.

Lemma 7 (All values are sums over atomic events)**.** *Let $(A_k, B_k)_{k=1}^{\mathcal{W}}$ be pairs of distributions over the universes $(\mathcal{U}_i)_{i=1}^{\mathcal{W}}$, let $f > 1$ and $n \in \mathbb{N}$ and let for all $k \in \{1,\dots,\mathcal{W}\}$ $\mathcal{B}(A_k, B_k, f_k, n) = (\mathcal{B}_k, \tilde{\ell}_k, \ell_k, f_k, 1)$ be bucket distributions (with error correction terms) and let T be a composition tree. Let $\varepsilon \geq 0$. Let*

$$(\mathcal{B}, \tilde{\ell}, \ell, f, u) := \prod_{k\in\{1,\dots,\mathcal{W}\}}^{T} \mathcal{B}(A_k, B_k, f_k, n),$$

Then, the following statements hold for all $i \in \{-n,\dots,n,\infty\}$:

- $\mathcal{B}(i) = \sum_{x \text{ s.t. } i_T(x)=i} \mathcal{B}(x)$
- $\tilde{\ell}(i) = \sum_{x \text{ s.t. } i_T(x)=i} \tilde{\ell}(x)$
- $\ell(i) = \sum_{x \text{ s.t. } i_T(x)=i} \ell(x)$

Proof. We show the lemma via structural induction over T. We only show the lemma for A, but the proof follows analogously for B.

If $T = \text{ⓟ}(A_i, B_i)$: Let $i \in \{-n, \dots, n, \infty\}$.

- By definition, $\mathcal{B}(x) = \Pr[x \leftarrow A]$ (c.f., Definition 12). Thus, $\mathcal{B}(i) = \sum_{x \text{ s.t. } i(x)=i} \Pr[x \leftarrow A] = \sum_{x \text{ s.t. } i(x)=i} \mathcal{B}(x)$.
- If $i \in \{-n, \dots, n\}$, then $\tilde{\ell}(i) = \sum_{x \text{ s.t. } i(x)=i} \Pr[x \leftarrow B] - \frac{\Pr[x \leftarrow A]}{f^i} = \sum_{x \text{ s.t. } i(x)=i} \tilde{\ell}(x)$. Otherwise $\tilde{\ell}(i) = 0 = \sum_{x \text{ s.t. } i(x)=i} 0 = \sum_{x \text{ s.t. } i(x)=i} \tilde{\ell}(x)$.
- If $i \in \{-n+1, \dots, n\}$, then $\ell(i) = \sum_{x \text{ s.t. } i(x)=i} \Pr[x \leftarrow B] - \frac{\Pr[x \leftarrow A]}{f^i} = \sum_{x \text{ s.t. } i(x)=i} \ell(x)$. Otherwise $\ell(i) = 0 = \sum_{x \text{ s.t. } i(x)=i} 0 = \sum_{x \text{ s.t. } i(x)=i} \ell(x)$.

If $T = T_1 \times T_2$: We assume the lemma holds for two bucket distributions $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ over a universe $\mathcal{U}_1$ and $(\mathcal{B}_2, \tilde{\ell}_2, \ell_2, f_2, u_2)$ over a universe $\mathcal{U}_2$ with composition trees T_1 and T_2. Then, with $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2$ and $T = T_1 \times T_2$, we have for $i \in \{-n+1, \dots, n\}$

$$
\begin{aligned}
\mathcal{B}_1 \times \mathcal{B}_2(i) &= \sum_{j,k \in \{-n,\dots,n\} \text{ s.t.} j+k=i} \mathcal{B}_1(j) \cdot \mathcal{B}_2(k) \\
&\overset{IV}{=} \sum_{j,k \in \{-n,\dots,n\} \text{ s.t. } j+k=i} \left(\sum_{x_1 \in \mathcal{U}_1 s.t. \ i_{T_1}(x_1)=j} \mathcal{B}_1(x_1) \right) \cdot \left(\sum_{x_2 \in \mathcal{U}_2 s.t. \ i_{T_2}(x_2)=k} \mathcal{B}_2(x_2) \right) \\
&= \sum_{x=(x_1,x_2) \in \mathcal{U}_1 \times \mathcal{U}_2 \text{ s.t. } i_{T_1}(x_1)+i_{T_2}(x_2)=i} \mathcal{B}_1(x_1) \cdot \mathcal{B}_2(x_2)
\end{aligned}
$$

We know from Definition 7 that $i_T(x) = i_{T_1}(x_1) + i_{T_2}(x_2)$, since $i_T(x) \in \{-n+1, \dots, n\}$.

$$
\begin{aligned}
&= \sum_{x=(x_1,x_2) \in \mathcal{U}_1 \times \mathcal{U}_2 \text{ s.t. } i_T(x)=i} \mathcal{B}_1(x_1) \cdot \mathcal{B}_2(x_2) \\
&= \sum_{x=(x_1,x_2) \in \mathcal{U} \text{ s.t. } i_T(x)=i} \mathcal{B}(x).
\end{aligned}
$$

For $i \in \{-n, \infty\}$ the proof follows analogously, where for $-n$ we have $j + k \leq -n$ and we know from Definition 7 that $i_T(x) = -n$ is equivalent to $i_{T_1}(x_1) + i_{T_2}(x_2) \leq -n$ and for ∞ we have $j + k > n$ and we know from Definition 7 that $i_T(x) = \infty$ is equivalent to $i_{T_1}(x_1) + i_{T_2}(x_2) \geq n$.

For the virtual error, we distinguish the following cases:

- $i_T(x) \in \{-n+1, \ldots, n\}$. Then,

$$
\begin{aligned}
&\tilde{\ell}_1 \times \tilde{\ell}_2(i) \\
&= \sum_{(k,l)\in\{-n,\ldots,n\}^2, k+l=i} \left(\frac{\mathcal{B}_1(k)}{f^k} + \tilde{\ell}_1(k)\right)\tilde{\ell}_2(l) + \tilde{\ell}_1(k)\left(\frac{\mathcal{B}_2(l)}{f^l} + \tilde{\ell}_2(l)\right) - \tilde{\ell}_1(k)\tilde{\ell}_2(l) \\
&= \sum_{(k,l)\in\{-n,\ldots,n\}^2, k+l=i} \frac{\mathcal{B}_1(k)}{f^k}\tilde{\ell}_2(l) + \tilde{\ell}_1(k)\frac{\mathcal{B}_2(l)}{f^l} + \tilde{\ell}_1(k)\tilde{\ell}_2(l) \\
&= \sum_{(k,l)\in\{-n,\ldots,n\}^2, k+l=i} \left(\frac{\sum_{x_1\in\mathcal{U}_1 s.t.\ i_{T_1}(x_1)=k} \mathcal{B}_1(x_1)}{f^k}\left(\sum_{x_2\in\mathcal{U}_2 s.t.\ i_{T_2}(x_2)=l} \tilde{\ell}_2(x_2)\right)\right. \\
&\quad + \left(\sum_{x_1\in\mathcal{U}_1 s.t.\ i_{T_1}(x_1)=k} \tilde{\ell}_1(x_1)\right)\frac{\sum_{x_2\in\mathcal{U}_2 s.t.\ i_{T_2}(x_2)=l} \mathcal{B}_2(l)}{f^l} \\
&\quad \left. + \left(\sum_{x_1\in\mathcal{U}_1 s.t.\ i_{T_1}(x_1)=k} \tilde{\ell}_1(x_1)\right)\left(\sum_{x_2\in\mathcal{U}_2 s.t.\ i_{T_2}(x_2)=l} \tilde{\ell}_2(x_2)\right)\right) \\
&= \sum_{(k,l)\in\{-n,\ldots,n\}^2, k+l=i} \sum_{x_1\in\mathcal{U}_1 s.t.\ i_{T_1}(x_1)=k} \sum_{x_2\in\mathcal{U}_2 s.t.\ i_{T_2}(x_2)=l} \left(\frac{\mathcal{B}_1(x_1)}{f^k}\tilde{\ell}_2(x_2)\right. \\
&\quad \left. + \tilde{\ell}_1(x_1)\frac{\mathcal{B}_2(l)}{f^l} + \tilde{\ell}_1(x_1)\tilde{\ell}_2(x_2)\right) \\
&= \sum_{(x_1,x_2)\in\mathcal{U}_1\times\mathcal{U}2 \ s.t.\ i_{T_1}(x_1)+i_{T_2}(x_2)=i} \left(\frac{\mathcal{B}_1(x_1)}{f^{i_{T_1}(x_1)}}\tilde{\ell}_2(x_2) + \tilde{\ell}_1(x_1)\frac{\mathcal{B}_2(l)}{f^{i_{T_2}(x_2)}} + \tilde{\ell}_1(x_1)\tilde{\ell}_2(x_2)\right)
\end{aligned}
$$

  We know from Definition 7 that $i_T(x) = i_{T_1}(x_1) + i_{T_2}(x_2)$, since $i_T(x) \in \{-n+1, \ldots, n\}$.

$$
= \sum_{x\in\mathcal{U} \ s.t.\ i_T(x)=i} \tilde{\ell}(x)
$$

- $i_T(x) = -n$. The proof of the case from above follows analogously with $k + l \leq -n$, since we know from Definition 7 that $i_T(x) = -n$ is equivalent to $i_{T_1}(x_1) + i_{T_2}(x_2) \leq -n$.
- $i_T(x) = \infty$.

$$
\begin{aligned}
&\tilde{\ell}_1 \times \tilde{\ell}_2(i) \\
= 0 = &\sum_{x\in\mathcal{U} \ s.t.\ i_T(x)=i} 0 \\
= &\sum_{x\in\mathcal{U} \ s.t.\ i_T(x)=i} \tilde{\ell}(x).
\end{aligned}
$$

If $T = \dagger T_1$:

We assume the lemma holds for two bucket distributions $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ over a universe $\mathcal{U}_1$ with a composition tree T_1. Then, with $\mathcal{U} = \mathcal{U}_1$ and $T = \dagger T_1$, we have for $i \in \{-n, \ldots, -n/2 - 1, n/2 + 1, \ldots, n\}$

$$
\dagger\mathcal{B}_1(i) = 0 = \sum_{x\in\emptyset} \mathcal{B}_1(x) = \sum_{x\in\mathcal{U} \text{ s.t. } i_T=i} \mathcal{B}_1(x)
$$

For $i = \infty$, we have

$$\dagger\mathcal{B}_1(\infty) = \mathcal{B}_1(\infty) \overset{\text{IH}}{=} \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x) = \infty} \mathcal{B}(x) = \sum_{x \in \mathcal{U} \text{ s.t. } i_T(x) = \infty} \mathcal{B}(x).$$

For $i \in \{-n/2 + 1, \ldots, n/2\}$ we have

$$\begin{aligned}
\dagger\mathcal{B}_1(i) &= \mathcal{B}_1(2i) + \mathcal{B}_1(2i-1) \\
&\overset{\text{IH}}{=} \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x) = 2i} \mathcal{B}_1(x) + \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x) = 2i-1} \mathcal{B}_1(x) \\
&= \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x) = 2i} \mathcal{B}(x) + \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x) = 2i-1} \mathcal{B}(x) \\
&= \sum_{x \in \mathcal{U} \text{ s.t. } i_T(x) = i} \mathcal{B}(x).
\end{aligned}$$

For $i = -n/2$ we have

$$\begin{aligned}
\dagger\mathcal{B}_1(-n/2) &= \mathcal{B}_1(-n) \\
&\overset{\text{IH}}{=} \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x) = -n} \mathcal{B}_1(x) \\
&= \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x) = -n} \mathcal{B}(x) \\
&= \sum_{x \in \mathcal{U} \text{ s.t. } i_T(x) = -n/2} \mathcal{B}(x).
\end{aligned}$$

We hence go forward to show the lemma for the error correction terms.

For the error correction terms and for $i \in \{-n, \ldots, -n/2 - 1, n/2 + 1, \ldots, n\}$

$$\dagger\tilde{\ell}_1(i) = 0 = \sum_{x \in \emptyset} \tilde{\ell}_1(x) = \sum_{x \in \mathcal{U} \text{ s.t. } i_T = i} \tilde{\ell}_1(x)$$

For $i = \infty$, we have

$$\dagger\tilde{\ell}_1(\infty) = 0 = \sum_{x \in \mathcal{U} \text{ s.t. } i_T(x) = \infty} 0 = \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x) = \infty} \tilde{\ell}(x) = \sum_{x \in \mathcal{U} \text{ s.t. } i_T(x) = \infty} \tilde{\ell}(x).$$

For $i \in \{-n/2+1, \dots, n/2\}$ we have

$$
\begin{aligned}
\dagger\tilde{\ell}_1(i) &= \tilde{\ell}_1(2i-1) + \mathcal{B}_1(2i-1)\left(\frac{1}{f^{2i-1}} - \frac{1}{f^{2i}}\right) + \tilde{\ell}_1(2i) \\
&\overset{\text{IH}}{=} \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x)=2i-1} \tilde{\ell}_1(x) + \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x)=2i-1} \mathcal{B}_1(x)\left(\frac{1}{f^{2i-1}} - \frac{1}{f^{2i}}\right) \\
&\quad + \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x)=2i} \tilde{\ell}_1(x) \\
&= \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x)=2i-1} \dagger\tilde{\ell}_1(x) - \mathcal{B}_1(x) \cdot \left(\frac{1}{f^{i_{T_1}(x)}} - \frac{1}{f^{2\cdot\lceil i_{T_1}(x)/2\rceil}}\right) \\
&\quad + \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x)=2i-1} \mathcal{B}_1(x)\left(\frac{1}{f^{2i-1}} - \frac{1}{f^{2i}}\right) \\
&\quad + \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x)=2i} \dagger\tilde{\ell}_1(x) - \mathcal{B}_1(x) \cdot \left(\frac{1}{f^{i_{T_1}(x)}} - \frac{1}{f^{2\cdot\lceil i_{T_1}(x)/2\rceil}}\right) \\
&= \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x)=2i-1} \dagger\tilde{\ell}_1(x) - \mathcal{B}_1(x) \cdot \left(\frac{1}{f^{2i-1}} - \frac{1}{f^{2i}}\right) + \mathcal{B}_1(x)\left(\frac{1}{f^{2i-1}} - \frac{1}{f^{2i}}\right) \\
&\quad + \sum_{x \in \mathcal{U} \text{ s.t. } i_{T_1}(x)=2i} \dagger\tilde{\ell}_1(x) \\
&= \sum_{x \in \mathcal{U} \text{ s.t. } i_T(x)=i} \dagger\tilde{\ell}_1(x)
\end{aligned}
$$

The proof for $\tilde{\ell}(i)$ in case $i = -n/2$ and the $\ell(i)$ follow analogously to the proof for $\tilde{\ell}(i)$ with the exception that the case $-n/2$ is analogous to the case ∞ instead to the cases $i \in \{-n+1, \dots, n\}$ for $\ell(i)$.

The proof for $\mathcal{B}_B$, $\tilde{\ell}_B$, and ℓ is symmetric. □

With Lemma 7 we now have a powerful tool for proving a set of properties for our error correction terms that will ultimately allow us to show the soundness of our results: We can relate every bucket value and every error correction term to the underlying events and can thus analyze our properties per event.

4.3 Helpful properties of error correction terms

In this rather technical subsection we present and show a set of helpful properties of our error correction terms that we require for our proof of soundness (and for our lower bound). We show that all error terms are positive (which means that not considering one of them can only increase the δ of our result), we show that our real error correction term is always smaller than the virtual error correction term and finally we show that for every event x, the virtual error correction term after an arbitrary amount of composition and squaring following the composition tree T still precisely captures $\Pr[x \leftarrow B] - \frac{\Pr[x \leftarrow A]}{f^{i_T}}$.

Lemma 8 (Positive real and virtual error correction terms)**.** *Let $(A_k, B_k)_{k=1}^{\mathcal{W}}$ be pairs of distributions over the universes $(\mathcal{U}_i)_{i=1}^{\mathcal{W}}$, let $f > 1$ and $n \in \mathbb{N}$ and let for all $k \in \{1, \dots, \mathcal{W}\}$ $\mathcal{B}(A_k, B_k, f_k, n) = (\mathcal{B}_k, \tilde{\ell}_k, \ell_k, f_k, 1)$ be bucket distributions (with error correction terms) and let T be a composition tree. Let $\varepsilon \geq 0$. Let*

$$
\mathcal{B}_T := (\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u) := \prod_{k \in \{1,\dots,\mathcal{W}\}}^{T} \mathcal{B}(A_k, B_k, f_k, n),
$$

The real error correction terms $\ell(i)$ and $\ell_B(i)$ for $i \in \{-n, \dots, n, \infty\}$ are positive, i.e., $\ell(i) \geq 0$ and $\ell_B(i) \geq 0$. Moreover, the virtual error correction terms $\tilde{\ell}(i)$ and $\tilde{\ell}_B(i)$ for $i \in \{-n, \dots, n, \infty\}$ are positive as well.

Proof. We show the lemma via structural induction over T.

For $T = \mathbb{P}(A, B)$, the real error correction term of an initial bucketing is calculated as the sum of error correction terms for each $x \in \mathcal{U}$ $\ell(x) = \Pr[x \leftarrow B] - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}}$ if $i_T(x) \notin \{n-, \infty\}$ and 0 otherwise. For $i_T(x) \in \{-n, \dots, n\}$ by definition we have $\Pr[x \leftarrow A] \leq f^{i_T(x)} \Pr[x \leftarrow B]$. Thus, for all $i \in \{-n, \dots, n, \infty\}$ are positive, i.e., $\ell(i) \geq 0$ and analogously we get $\ell_B(i) \geq 0$.

For $T = T_1 \times T_2$, $\mathcal{B}_T$ is the result of composing two bucket distributions $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ and $(\mathcal{B}_2, \tilde{\ell}_2, \ell_2, f_2, u_2)$. By induction hypothesis, ℓ_1 and ℓ_2 are positive. We calculate the composed error correction terms as either 0 (if $i \in \{-n, \infty\}$) or as

$$\ell(i) = \ell_{A_1 \times A_2}(i) = \sum_{j,k \text{ s.t. } j+k=i} \left(\left(\frac{\mathcal{B}_{A_1}(j)}{f^j} \right) \ell_2(k) + \left(\frac{\mathcal{B}_{A_2}(k)}{f^k} \right) \ell_1(j) + \ell_1(j) \tilde{\ell}_2(k) \right),$$

which is positive as well since all the error correction terms and all bucket terms are positive.

For $T = \dagger T_1$, We calculate the error correction terms as either 0 (if $i \in \{-n, \dots, -n/2 - 1, n/2 + 1, \dots, n, \infty\}$) or as

$$\ell(i) = \dagger \ell_1(i) = \ell_1(2i-1) + \mathcal{B}_1(2i-1) \left(\frac{1}{f^{2i-1}} - \frac{1}{f^{2i}} \right) + \ell_1(2i),$$

which is positive as well since all the error correction terms and all bucket terms are positive. [4] Analogously, we can show that the virtual error correction terms $\tilde{\ell}$ are positive as well. □

We now show that the real error correction term is smaller than the virtual error correction term.

Lemma 9 (The real error ℓ is smaller than the virtual error $\tilde{\ell}$)**.** *Let $(A_k, B_k)_{k=1}^{\mathcal{W}}$ be pairs of distributions over the universes $(\mathcal{U}_i)_{i=1}^{\mathcal{W}}$, let $f > 1$ and $n \in \mathbb{N}$ and let for all $k \in \{1, \dots, \mathcal{W}\}$ $\mathcal{B}(A_k, B_k, f_k, n) = (\mathcal{B}_k, \tilde{\ell}_k, \ell_k, f_k, 1)$ be bucket distributions (with error correction terms) and let T be a composition tree. Let $\varepsilon \geq 0$. Let*

$$\mathcal{B}_T := (\mathcal{B}, \tilde{\ell}, \ell, f, u) := \prod_{k \in \{1, \dots, \mathcal{W}\}}^{T} \mathcal{B}(A_k, B_k, f_k, n),$$

Then, the real error is always smaller than the virtual error: $\ell(x) \leq \tilde{\ell}(x)$ and $\ell_B(x) \leq \tilde{\ell}_B(x)$.

Proof. We show the lemma via structural induction over T.

For $T = \mathbb{P}(A, B)$: We know that $\tilde{\ell}(x) \geq 0$. By definition, for $u = 1$, either $\ell(x) = 0$ or $\ell_T(x) = \tilde{\ell}_T(x)$ holds. Thus, $\ell_T(x) \leq \tilde{\ell}_T(x)$.

For $T = T_1 \times T_2$: $\mathcal{B}_T$ is the result of composing two bucket distributions $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ and $(\mathcal{B}_2, \tilde{\ell}_2, \ell_2, f_2, u_2)$. By induction hypothesis, $\ell_1 \leq \tilde{\ell}_1$ and $\ell_2 \leq \tilde{\ell}_2$. For $i_T(x) = -n$, $\ell(x) = 0$. By Lemma 8 we know that $0 \leq \tilde{\ell}(x)$, hence $\ell(x) = 0 \leq \tilde{\ell}(x)$. For $i_T(x) \neq -n$, with $x_1 \in \mathcal{U}_1$ and $x_2 \in \mathcal{U}_2$ we have

$$\begin{aligned}
\ell(x) = \ell_1 \times \ell_2(x) &= \left(\frac{\Pr[x_1 \leftarrow A_1]}{f^{i_{T_1}(x_1)}} + \ell_1(x_1) \right) \ell_2(x_2) + \left(\frac{\Pr[x_2 \leftarrow A_2]}{f^{i_{T_2}(x_2)}} + \ell_2(x_2) \right) \ell_1(x_1) - \ell_1(x_1)\ell_2(x_2) \\
&= \left(\frac{\Pr[x_1 \leftarrow A_1]}{f^{i_{T_1}(x_1)}} \right) \underbrace{\ell_2(x_2)}_{\overset{IH}{\leq} \tilde{\ell}_2(x_2)} + \left(\frac{\Pr[x_2 \leftarrow A_2]}{f^{i_{T_2}(x_2)}} \right) \underbrace{\ell_1(x_1)}_{\overset{IH}{\leq} \tilde{\ell}_1(x_1)} + \underbrace{\ell_1(x_1)}_{\overset{IH}{\leq} \tilde{\ell}_1(x_1)} \underbrace{\ell_2(x_2)}_{\overset{IH}{\leq} \tilde{\ell}_2(x_2)} \\
&\overset{IH}{\leq} \left(\frac{\Pr[x_1 \leftarrow A_1]}{f^{i_{T_1}(x_1)}} \right) \tilde{\ell}_2(x_2) + \left(\frac{\Pr[x_2 \leftarrow A_2]}{f^{i_{T_2}(x_2)}} \right) \tilde{\ell}_1(x_1) + \tilde{\ell}_1(x_1)\tilde{\ell}_2(x_2) \\
&= \tilde{\ell}_1 \times \tilde{\ell}_2(x) = \tilde{\ell}(x)
\end{aligned}$$

[4] Note that in the case $-n/2$ there is only one term instead of two. This term, however, is still positive.

For $T = \dagger T_1$: This case directly holds by induction hypothesis, as the squaring operation is analogously defined for the real and the virtual error.

□

We now show our main lemma for the lower bound on δ: the virtual error correction term is precise for any event with an index other than ∞. We can directly use this lemma to get a lower bound for δ if we ignore the bucket with index ∞. Note that although the virtual error is precise on a per-event basis, events can still be misplaced and thus negatively contribute to δ if we use the virtual error correction term. For our upper bound on δ we circumvent this problem by over-approximating misplaced events (using the real error correction term) and by not using error correction terms in some buckets with a bucket factor f^i close to e^ε.

Lemma 10 (Characterizing the virtual error after compositions and rescaling)**.** *Let $(A_k, B_k)_{k=1}^{\mathcal{W}}$ be pairs of distributions over the universes $(\mathcal{U}_i)_{i=1}^{\mathcal{W}}$, let $f > 1$ be the bucketing factor of the root node and $n \in \mathbb{N}$ and let for all $k \in \{1, \dots, \mathcal{W}\}$ $\mathcal{B}(A_k, B_k, f_k, n) = (\mathcal{B}_k, \tilde{\ell}_k, \ell_k, f_k, 1)$ be bucket distributions (with error correction terms) and let T be a composition tree. Let $\varepsilon \geq 0$. Let*

$$\mathcal{B}_T := (\mathcal{B}, \tilde{\ell}, \ell, f, u) := \prod_{k \in \{1,\dots,\mathcal{W}\}}^{T} \mathcal{B}(A_k, B_k, f_k, n),$$

Then, for all $i \in \{-n, \dots, n\}$ we have

$$\tilde{\ell}(x) = \Pr[x \leftarrow B] - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}}$$

Proof. We show the lemma via structural induction over T. For $T = \mathbb{P}(A_1, B_1)$, the statement follows by construction:

$$\Pr[x \leftarrow B] - \frac{\Pr[x \leftarrow A]}{f^i} = \Pr[x \leftarrow B_1] - \frac{\Pr[x \leftarrow A_1]}{f^i} = \tilde{\ell}(i),$$

where f is the bucketing factor of the leaf.

For $T = T_1 \times T_2$, $\mathcal{B}_T$ is the result of composing two bucket distributions $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ and $(\mathcal{B}_2, \tilde{\ell}_2, \ell_2, f_2, u_2)$, both with the same bucketing factor f as the composition node. By induction hypothesis, the statement holds for $\tilde{\ell}_1$ and $\tilde{\ell}_2$. By definition of the error correction term composition we get with $x_1 \in \mathcal{U}_1$

and $x_2 \in \mathcal{U}_2$

$$
\begin{aligned}
\tilde{\ell}(x) =& (\tilde{\ell}_1 \times \tilde{\ell}_2)(x) \\
=& \left(\frac{\Pr\left[x_1 \leftarrow A_1\right]}{f^{i_{T_1}(x_1)}} + \tilde{\ell}_1(x_1)\right)\tilde{\ell}_2(x_2) + \left(\frac{\Pr\left[x_2 \leftarrow A_2\right]}{f^{i_{T_2}(x_2)}} + \tilde{\ell}_2(x_2)\right)\tilde{\ell}_1(x_1) - \tilde{\ell}_1(x_1)\tilde{\ell}_2(x_2) \\
=& \frac{\Pr\left[x_1 \leftarrow A_1\right]}{f^{i_{T_1}(x_1)}} \cdot \tilde{\ell}_2(x_2) + \frac{\Pr\left[x_2 \leftarrow A_2\right]}{f^{i_{T_2}(x_2)}} \cdot \tilde{\ell}_1(x_1) + \tilde{\ell}_1(x_1)\tilde{\ell}_2(x_2) \\
\overset{IH}{=}& \frac{\Pr\left[x_1 \leftarrow A_1\right]}{f^{i_{T_1}(x_1)}} \cdot \left(\Pr\left[x_2 \leftarrow B_2\right] - \frac{\Pr\left[x_2 \leftarrow A_2\right]}{f^{i_{T_2}(x_2)}}\right) \\
&+ \frac{\Pr\left[x_2 \leftarrow A_2\right]}{f^{i_{T_2}(x_2)}} \cdot \left(\Pr\left[x_1 \leftarrow B_1\right] - \frac{\Pr\left[x_1 \leftarrow A_1\right]}{f^{i_{T_1}(x_1)}}\right) \\
&+ \left(\Pr\left[x_1 \leftarrow B_1\right] - \frac{\Pr\left[x_1 \leftarrow A_1\right]}{f^{i_{T_1}(x_1)}}\right) \cdot \left(\Pr\left[x_2 \leftarrow B_2\right] - \frac{\Pr\left[x_2 \leftarrow A_2\right]}{f^{i_{T_2}(x_2)}}\right) \\
=& \frac{\Pr\left[x_1 \leftarrow A_1\right]}{f^{i_{T_1}(x_1)}} \cdot \Pr\left[x_2 \leftarrow B_2\right] - \frac{\Pr\left[x \leftarrow A\right]}{f^{i_T(x)}} \\
&+ \frac{\Pr\left[x_2 \leftarrow A_2\right]}{f^{i_{T_2}(x_2)}} \cdot \Pr\left[x_1 \leftarrow B_1\right] - \frac{\Pr\left[x \leftarrow A\right]}{f^{i_T(x)}} \\
&+ \Pr\left[x \leftarrow B\right] - \frac{\Pr\left[x_1 \leftarrow A_1\right]}{f^{i_{T_1}(x_1)}} \cdot \Pr\left[x_2 \leftarrow B_2\right] - \frac{\Pr\left[x_2 \leftarrow A_2\right]}{f^{i_{T_2}(x_2)}} \cdot \Pr\left[x_1 \leftarrow B_1\right] + \frac{\Pr\left[x \leftarrow A\right]}{f^{i_T(x)}} \\
=& \Pr\left[x \leftarrow B\right] - \frac{\Pr\left[x \leftarrow A\right]}{f^{i_T(x)}}
\end{aligned}
$$

For $T = \dagger T_1$, we know that for all $x \in \mathcal{U}$, $i_{T_1}(x) \in \{-n/2, \ldots, n/2\} \cup \{\infty\}$. Since the index ∞ is excluded in our lemma, we focus on the remaining values for the index. Note that the bucketing factor in this case changes from f (of the child node) to f^2 (of the squaring node). By induction hypothesis, we have

$$\tilde{\ell}_1(x) = \Pr\left[x \leftarrow B\right] - \frac{\Pr\left[x \leftarrow A_1\right]}{f^{i_T(x)}}$$

Consequently and since $i_{T_1}(x) \in \{-n/2, \ldots, n/2\}$, we get,

$$
\begin{aligned}
\tilde{\ell}(x) = \dagger\tilde{\ell}_1(x) &= \tilde{\ell}_1(x) + \mathcal{B}_{A_1}(x) \cdot \left(\frac{1}{f^{i_{T_1}(x)}} - \frac{1}{f^{2 \cdot \lceil i_{T_1}(x)/2 \rceil}}\right) \\
&\overset{IH}{=} \Pr\left[x \leftarrow B\right] - \frac{\Pr\left[x \leftarrow A\right]}{f^{i_{T_1}(x)}} + \Pr\left[x \leftarrow A\right] \cdot \left(\frac{1}{f^{i_{T_1}(x)}} - \frac{1}{f^{2 \cdot \lceil i_{T_1}(x)/2 \rceil}}\right) \\
&= \Pr\left[x \leftarrow B\right] - \frac{\Pr\left[x \leftarrow A\right]}{f^{i_{T_1}(x)}} + \Pr\left[x \leftarrow A\right] \cdot \left(\frac{1}{f^{i_{T_1}(x)}} - \frac{1}{f^{2 i_T(x)}}\right) \\
&= \Pr\left[x \leftarrow B\right] - \frac{\Pr\left[x \leftarrow A\right]}{(f^2)^{i_T(x)}}.
\end{aligned}
$$

□

4.4 The approximated delta with error correction

Finally, we define how to calculate a sound upper bound on δ based on a bucket distribution with error correction terms. We note that when using the real error correction term, events cannot harm the soundness by being misplaced as a result of parts of the event having been placed in the smallest bucket (with index $-n$). However, every composition can misplace an arbitrary event by one bucket. This slight misplacement poses a problem for a small number of buckets with a bucket factor very close to e^{ε}, as they can now contain events

that actually have a negative contribution to δ: $\Pr[x \leftarrow A] - e^{\varepsilon}\Pr[x \leftarrow B] < 0$. Every bucket distribution carries a value u that increases by 1 for every composition (and that can be reduced by squaring). If j_ε is the index of the bucket with the smallest bucket factor larger than e^ε, we don't consider the the error correction term for buckets with index $i < j_\varepsilon + u$ and instead fall back to the definition from Definition 4 for those buckets. For the remaining buckets *withi* $\geq j_\varepsilon + u$, which typically is the vast majority of buckets, we make use of the real error correction term to reduce the error.

Definition 14 (Approximated delta with error correction)**.** *Let $(A_k, B_k)_{k=1}^{\mathcal{W}}$ be pairs of distributions over the universes $(\mathcal{U}_i)_{i=1}^{\mathcal{W}}$, let $\varepsilon > 0$ and $n \in \mathbb{N}$ and let for all $k \in \{1, \ldots, \mathcal{W}\}$ $\mathcal{B}(A_k, B_k, f_k, n) = (\mathcal{B}_k, \tilde{\ell}_k, \ell_k, f_k, 1)$ be bucket distributions (with error correction terms) and let T be a composition tree. Let $\varepsilon \geq 0$. Let*

$$\mathcal{B}_T := (\mathcal{B}, \tilde{\ell}, \ell, f, u) := \prod_{k \in \{1,\ldots,\mathcal{W}\}}^{T} \mathcal{B}(A_k, B_k, f_k, n),$$

We define $\delta(\mathcal{B}_T, \varepsilon)$ with $j_\varepsilon \in \mathbb{N}$ such that $f^{j_\varepsilon - 1} < e^\varepsilon \leq f^{j_\varepsilon}$ as

$$\begin{aligned}\delta(\mathcal{B}_T, \varepsilon) :=& \\ &\sum_{i \in \{j_\varepsilon, \ldots, j_\varepsilon + u - 1\}} \mathcal{B}(i) - \frac{e^\varepsilon \mathcal{B}(i)}{f^i} \\ &+ \sum_{i \in \{j_\varepsilon + u, \ldots, n\}} \left(\mathcal{B}(i) - e^\varepsilon \left(\frac{\mathcal{B}(i)}{f^i} + \ell(i)\right)\right) + \mathcal{B}(\infty)\end{aligned}$$

Moreover, for all individual events $x \leftarrow \mathcal{U}$ we define

$$\delta(\mathcal{B}_T, x, \varepsilon) := \begin{cases} \Pr[x \leftarrow A] \cdot \left(1 - \frac{e^\varepsilon}{f^{i_T(x)}}\right) & 1.\ \text{if } j_\varepsilon \leq i_T(x) \leq j_\varepsilon + u - 1 \\ \Pr[x \leftarrow A] - e^\varepsilon \left(\frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} + \ell_E(x)\right) & 2.\ \text{if } j_\varepsilon + u \leq i_T(x) \leq n \\ \Pr[x \leftarrow A] & 3.\ \text{if } i_T(x) = \infty \\ 0 & 4.\ \text{otherwise} \end{cases}$$

Note that if $j > n$, we only consider elements in the bucket B_∞.

Next we show that the real error correction terms are bounded by the value of u: For every event x the real error correction term $\ell(x)$ can never exceed a fraction of $\frac{1}{f^{i_T(x)-u}} - \frac{1}{f^{i_T(x)}}$ of the probability of the event. Intuitively, this means that the value of the real error correction term can never be larger than what a *misplacement by u buckets* would result in.

Lemma 11 (An upper bound for ℓ)**.** *Let $(A_k, B_k)_{k=1}^{\mathcal{W}}$ be pairs of distributions over the universes $(\mathcal{U}_i)_{i=1}^{\mathcal{W}}$, let $f > 1$ and $n \in \mathbb{N}$ and let for all $k \in \{1, \ldots, \mathcal{W}\}$ $\mathcal{B}(A_k, B_k, f_k, n) = (\mathcal{B}_k, \tilde{\ell}_k, \ell_k, f_k, 1)$ be bucket distributions (with error correction terms) and let T be a composition tree. Let $\varepsilon \geq 0$ and with $j_\varepsilon \in \mathbb{N}$ such that $f^{j_\varepsilon - 1} < e^\varepsilon \leq f^{j_\varepsilon}$. Let*

$$\mathcal{B}_T := (\mathcal{B}, \tilde{\ell}, \ell, f, u) := \prod_{k \in \{1,\ldots,\mathcal{W}\}}^{T} \mathcal{B}(A_k, B_k, f_k, n),$$

If $j_\varepsilon + u \leq i_T(x) \neq \infty$, then the error correction term never makes a negative contribution to the approximated delta with error correction:

$$\ell(x) \leq \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-u}} - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}}$$

Proof. We show the lemma via structural induction over T.

Let $T = \mathbin{\text{🪐}}(A, B)$. If $i_T(x) = -n$ then

$$\ell(x) = 0 \leq \Pr[x \leftarrow A] \cdot \left(\frac{1}{f^{-n-1}} - \frac{1}{f^{-n}}\right).$$

Otherwise, if $i_T(x) > -n$, we know that by definition of $i_T(x)$ we have $f^{i_T(x)-1}\Pr[x \leftarrow B] \leq \Pr[x \leftarrow A]$

$$\begin{aligned}\ell(x) =& \Pr[x \leftarrow B] - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} \\ \leq& \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-1}} - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}}.\end{aligned}$$

Let $T = T_1 \times T_2$. If $i_T(x) = -n$, then $\ell(x) = 0 \leq \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-u}} - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}}$. Otherwise, $\mathcal{B}_T$ is the result of composing two bucket distributions $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ and $(\mathcal{B}_2, \tilde{\ell}_2, \ell_2, f_2, u_2)$. By induction hypothesis, the statement holds for ℓ_1 ℓ_2. For $x_1 \in \mathcal{U}_1$ and $x_2 \in \mathcal{U}_2$ we know that $i_T(x) = i_{T_1}(x_1) + i_{T_2}(x_2)$. Moreover, we know that $\Pr[x \leftarrow A] = \Pr[x_1 \leftarrow A_1] \cdot \Pr[x_2 \leftarrow A_2]$. Thus, for $u = u_1 + u_2$ we get

$$\begin{aligned}\ell(x) &= \left(\frac{\Pr[x_1 \leftarrow A_1]}{f^{i_{T_1}(x_1)}} + \ell_1(x_1)\right)\ell_2(x_2) + \left(\frac{\Pr[x_2 \leftarrow A_2]}{f^{i_{T_2}(x_2)}} + \ell_2(x_2)\right)\ell_1(x_1) - \ell_1(x_1)\ell_2(x_2) \\ &= \left(\frac{\Pr[x_1 \leftarrow A_1]}{f^{i_{T_1}(x_1)}}\right)\ell_2(x_2) + \left(\frac{\Pr[x_2 \leftarrow A_2]}{f^{i_{T_2}(x_2)}}\right)\ell_1(x_1) + \ell_1(x_1)\ell_2(x_2)\end{aligned}$$

$$\begin{aligned}&\overset{IH}{\leq} \left(\frac{\Pr[x_1 \leftarrow A_1]}{f^{i_{T_1}(x_1)}}\right)\left(\frac{\Pr[x_2 \leftarrow A_2]}{f^{i_{T_2}(x_2)-(u-u_1)}} - \frac{\Pr[x_2 \leftarrow A_2]}{f^{i_{T_2}(x_2)}}\right) \\ &\quad + \left(\frac{\Pr[x_2 \leftarrow A_2]}{f^{i_{T_2}(x_2)}}\right)\left(\frac{\Pr[x_1 \leftarrow A_1]}{f^{i_{T_1}(x_1)-u_1}} - \frac{\Pr[x_1 \leftarrow A_1]}{f^{i_{T_1}(x_1)}}\right) \\ &\quad + \left(\frac{\Pr[x_1 \leftarrow A_1]}{f^{i_{T_1}(x_1)-u_1}} - \frac{\Pr[x_1 \leftarrow A_1]}{f^{i_{T_1}(x_1)}}\right)\left(\frac{\Pr[x_2 \leftarrow A_2]}{f^{i_{T_2}(x_2)-(u-u_1)}} - \frac{\Pr[x_2 \leftarrow A_2]}{f^{i_{T_2}(x_2)}}\right) \\ &= \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-(u-u_1)}} - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} \\ &\quad + \frac{\Pr[x \leftarrow A]}{f^{i_{T_2}(x_2)+i_{T_1}(x_1)-u_1}} - \frac{\Pr[x \leftarrow A]}{f^{i_{T_2}(x_2)+i_{T_1}(x_1)}} \\ &\quad + \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-u_1-(u-u_1)}} - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-(u-u_1)}} - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-u_1}} + \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} \\ &= \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-u_2}} - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} \\ &\quad + \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-u_1}} - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} \\ &\quad + \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-u}} - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-u_2}} - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-u_1}} + \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} \\ &= \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-u}} - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}}\end{aligned}$$

Let $T = \dagger T_1$. In this case, if the child node has a bucketing factor of f_1 and a value of u_1, the squaring node has a bucketing factor of $f_1^2 = f$ and a value of $u = \lceil u_1/2 \rceil + 1$. We know that $\dagger\ell(x) = \ell_1(x) + \mathcal{B}_1(x) \cdot \left(\frac{1}{f_1^{i_{T_1}(x)}} - \frac{1}{f_1^{2 \cdot \lceil i_{T_1}(x)/2 \rceil}}\right)$. Since we excluded $i_T = \infty = i_{T_1}$ and $j_\varepsilon + u \leq i_T$, we know that $i_T \in \{0, \ldots, n/2\}$.

Thus,

$$\begin{aligned}
\ell(x) &= \dagger\ell_1(x) \\
&= \ell_1(x) + \mathcal{B}_1(x) \cdot \left(\frac{1}{f_1^{i_{T_1}(x)}} - \frac{1}{f_1^{2 \cdot \lceil i_{T_1}(x)/2 \rceil}} \right) \\
&\overset{IH}{\leq} \frac{\Pr[x \leftarrow A]}{f_1^{i_{T_1}(x)-u_1}} - \frac{\Pr[x \leftarrow A]}{f_1^{i_{T_1}(x)}} + \mathcal{B}_1(x) \cdot \left(\frac{1}{f_1^{i_{T_1}(x)}} - \frac{1}{f_1^{2 \cdot \lceil i_{T_1}(x)/2 \rceil}} \right) \\
&= \frac{\Pr[x \leftarrow A]}{f_1^{i_{T_1}(x)-u_1}} - \frac{\Pr[x \leftarrow A]}{f_1^{i_{T_1}(x)}} + \frac{\Pr[x \leftarrow A]}{f_1^{i_{T_1}(x)}} - \frac{\Pr[x \leftarrow A]}{f_1^{2 \cdot \lceil i_{T_1}(x)/2 \rceil}} \\
&= \frac{\Pr[x \leftarrow A]}{(f_1^2)^{\frac{i_{T_1}(x)-u_1}{2}}} - \frac{\Pr[x \leftarrow A]}{(f_1^2)^{i_T(x)}} \\
&\leq \frac{\Pr[x \leftarrow A]}{(f_1^2)^{\lceil i_{T_1}(x)/2 \rceil - (\lceil u_1/2 \rceil + 1)}} - \frac{\Pr[x \leftarrow A]}{(f_1^2)^{i_T(x)}} \\
&= \frac{\Pr[x \leftarrow A]}{(f_1^2)^{i_T(x)-u}} - \frac{\Pr[x \leftarrow A]}{(f_1^2)^{i_T(x)}}
\end{aligned}$$

□

From Lemma 11 we can deduct that no event in a bucket with index $i \geq j_\varepsilon + u$ can have a negative impact on δ. Since moreover for each event we consider an impact that is at least as large as the actual impact of the event (as in the precise calculation of δ from Lemma 1) we can show the soundness of our result:

Lemma 12 (Soundness of the approximated delta with error correction)**.** *Let $(A_k, B_k)_{k=1}^{\mathcal{W}}$ be pairs of distributions over the universes $(\mathcal{U}_i)_{i=1}^{\mathcal{W}}$, let $f > 1$ and $n \in \mathbb{N}$ and let for all $k \in \{1, \ldots, \mathcal{W}\}$ $\mathcal{B}(A_k, B_k, f_k, n) = (\mathcal{B}_k, \tilde{\ell}_k, \ell_k, f_k, 1)$ be bucket distributions (with error correction terms) and let T be a composition tree. Let $\varepsilon \geq 0$. Let*

$$\mathcal{B}_T := (\mathcal{B}, \tilde{\ell}, \ell, f, u) := \prod_{k \in \{1,\ldots,\mathcal{W}\}}^{T} \mathcal{B}(A_k, B_k, f_k, n),$$

Then, the following statement holds:

$$\delta(\mathcal{B}_T, \varepsilon) \geq \sum_{x \in \mathcal{U}} \max\left(0, \Pr[x \leftarrow A] - e^\varepsilon \Pr[x \leftarrow B]\right)$$

Proof. Let

$$S_i = \{x \in \mathcal{U} \text{ s.t. } i_T(x) = i\}$$

As these S_i define a partitioning of $\mathcal{U}$, the definition these S_i implies $\bigcup_{i \in \{-n,\ldots,n,\infty\}} S_i = \mathcal{U}$.

As $\delta(\mathcal{B}_T, \varepsilon)$ is is a sum over $\mathcal{B}_A$ and ℓ, Lemma 7 implies that

$$\delta(\mathcal{B}_T, \varepsilon) = \sum_{x \in \mathcal{U}} \delta(\mathcal{B}_T, x, \varepsilon)$$

We next distinguish the the four cases of the definition of $\delta(\mathcal{B}_T, x, \varepsilon)$.

Case 1. This case occurs if $j_\varepsilon \leq i_T(x) \leq j_\varepsilon + u - 1$. By Lemma 6, we know the following

$$\begin{aligned}
\Pr[x \leftarrow A] \leq & f^{i_T(x)} \Pr[x \leftarrow B] \\
\Leftrightarrow \quad \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} \leq & \Pr[x \leftarrow B] \\
\Leftrightarrow \quad \Pr[x \leftarrow A] - e^\varepsilon \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} \geq & \Pr[x \leftarrow A] - e^\varepsilon \Pr[x \leftarrow B]
\end{aligned}$$

By definition of $\delta(\mathcal{B}_T, \varepsilon)$, we get

$$\delta(\mathcal{B}_T, x, \varepsilon) = \Pr[x \leftarrow A] - e^{\varepsilon}\frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} \geq \Pr[x \leftarrow A] - e^{\varepsilon}\Pr[x \leftarrow B]$$

Moreover, as $i_T(x) >= j_\varepsilon$, we know that $e^{\varepsilon} \leq f^{i_T(x)}$. Hence, we also get

$$\delta(\mathcal{B}_T, x, \varepsilon) = \Pr[x \leftarrow A] - \underbrace{\frac{e^{\varepsilon}}{f^{i_T(x)}}}_{\leq 1}\Pr[x \leftarrow A] \geq 0$$

Case 2. This case occurs if $i_T(x) \geq j_\varepsilon + u$.

We show two things: (i)

$$\Pr[x \leftarrow A] - e^{\varepsilon}\left(\frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} + \ell(x)\right)$$

by Lemma 11 we know that $\ell(x) \leq \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-u}} - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}}$ holds; hence, we get

$$\begin{aligned}
&\geq \Pr[x \leftarrow A] - e^{\varepsilon}\left(\frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} + \frac{\Pr[x \leftarrow A]}{f^{i_T(x)-u}} - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}}\right)\\
&= \Pr[x \leftarrow A] - e^{\varepsilon}\left(\frac{\Pr[x \leftarrow A]}{f^{i_T(x)-u}}\right)\\
&\geq \Pr[x \leftarrow A] - \frac{f^{j_\varepsilon}}{f^{i_T(x)-u}}\Pr[x \leftarrow A]\\
&= \Pr[x \leftarrow A] \cdot \left(1 - \frac{f^{j_\varepsilon}}{f^{i_T(x)-u}}\right)
\end{aligned}$$

as by assumption $i_T(x) \geq j_\varepsilon + u$, we get

$$\geq 0$$

(ii) Note that

$$\frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} + \underbrace{\ell(x)}_{\leq \tilde{\ell}(x)} \leq \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} + \tilde{\ell}(x) \overset{Lemma\ 10}{=} \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} + \Pr[x \leftarrow B] - \frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} = \Pr[x \leftarrow B]$$

Thus,

$$\Pr[x \leftarrow A] - e^{\varepsilon}\left(\frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} + \ell(x)\right) \geq \Pr[x \leftarrow A] - e^{\varepsilon}\Pr[x \leftarrow B]$$

From (i) and (ii) we get

$$\begin{aligned}
\delta(A, B, x, f, n, u, \varepsilon) &= \Pr[x \leftarrow A] - e^{\varepsilon}\left(\frac{\Pr[x \leftarrow A]}{f^{i_T(x)}} + \ell(x)\right)\\
&\geq \max(0, \Pr[x \leftarrow A] - e^{\varepsilon}\Pr[x \leftarrow B])
\end{aligned}$$

Case 3. By definition of δ, we have $\delta(x) = \Pr[x \leftarrow A] > \max(0, \Pr[x \leftarrow A] - e^{\varepsilon}\Pr[x \leftarrow B])$.

Case 4. Thus, for all x with $i_T(x) \leq j_\varepsilon$,

$$\begin{aligned}
&\Pr[x \leftarrow A] - e^{\varepsilon}\Pr[x \leftarrow B]\\
\leq &\Pr[x \leftarrow A] - f^{j_\varepsilon}\Pr[x \leftarrow B]\\
\leq &\Pr[x \leftarrow A] - f^{i_T(x)}\Pr[x \leftarrow B] \overset{\text{Lemma 6}}{\leq} 0
\end{aligned}$$

and thus,

$$\delta(x) = 0 = \max\left(0, \Pr\left[x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \leftarrow B\right]\right)$$

□

We now present our main result: Given any value for $\varepsilon \geq 0$ and a value δ_ε, s.t. the distributions are tightly $(\varepsilon, \delta_\varepsilon)$-differentially private, the value δ calculated as in Definition 14 presents a sound upper bound on δ_ε from Lemma 1 and we introduce a lower bound δ^{low}, s.t. δ^{low} presents a lower bound on δ_ε.

Theorem 2 (Buckets with error correction terms are sound)**.** *Let $(A_k, B_k)_{k=1}^{\mathcal{W}}$ be pairs of distributions over the universes $\mathcal{U} := (\mathcal{U}_i)_{i=1}^{\mathcal{W}}$, let $f > 1$ and $n \in \mathbb{N}$ and let for all $k \in \{1, \ldots, \mathcal{W}\}$ $\mathcal{B}(A_k, B_k, f_k, n) = (\mathcal{B}_k, \tilde{\ell}_k, \ell_k, f_k, 1)$ be bucket distributions (with error correction terms) and let T be a composition tree. Let $\varepsilon \geq 0$ and $j_\varepsilon \in \mathbb{N}$ s.t. $f^{j_\varepsilon - 1} < e^{\varepsilon} \leq f^{j_\varepsilon}$,*

$$\mathcal{B}_T := (\mathcal{B}, \tilde{\ell}, \ell, f, u) := \prod_{k \in \{1,\ldots,\mathcal{W}\}}^{T} \mathcal{B}(A_k, B_k, f_k, n),$$

$$\delta_\varepsilon = \max\left(\sum_{x \in \mathcal{U}} \max\left(\Pr\left[x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \leftarrow B\right], 0\right), \sum_{x \in \mathcal{U}} \max\left(\Pr\left[x \leftarrow B\right] - e^{\varepsilon}\Pr\left[x \leftarrow A\right], 0\right)\right)$$

$$\delta^{\text{low}} := \sum_{i \in \{j_\varepsilon, \ldots, n\}} \max\left(0, \mathcal{B}(i) - e^{\varepsilon}\left(\frac{\mathcal{B}(i)}{f^i} + \tilde{\ell}(i)\right)\right)$$

Then, $\prod_{k=1}^{\mathcal{W}} A_k$ and $\prod_{k=1}^{\mathcal{W}} B_k$ are $(\varepsilon, \delta_\varepsilon)$-differentially private, and

$$\delta^{\text{low}} \leq \delta_\varepsilon \leq \delta(\mathcal{B}_T, \varepsilon),$$

Proof. Lemma 1 shows that $\prod_{k=1}^{\mathcal{W}} A_k$ and $\prod_{k=1}^{\mathcal{W}} B_k$ are $(\varepsilon, \delta_\varepsilon)$-differentially private, and Lemma 12 proves that $\delta_\varepsilon \leq \delta(\mathcal{B}_T, \varepsilon)$ holds true.

Next, we show that $\delta^{\text{low}} \leq \delta_\varepsilon$:

$$\begin{aligned}
\delta^{\text{low}} &= \sum_{i \in \{j_\varepsilon, \ldots, n\}} \max\left(0, \mathcal{B}(i) - e^{\varepsilon}\left(\frac{\mathcal{B}(i)}{f^i} + \tilde{\ell}(i)\right)\right) \\
&\overset{\text{Lemma 7}}{=} \sum_{i \in \{j_\varepsilon, \ldots, n\}} \max\left(0, \sum_{x \in \mathcal{U}, i_T(x) = i} \Pr\left[x \leftarrow A\right] - e^{\varepsilon}\left(\frac{\Pr\left[x \leftarrow A\right]}{f^i} + \tilde{\ell}(x)\right)\right) \\
&\overset{\text{Lemma 10}}{=} \sum_{i \in \{-n, \ldots, n, \infty\}} \max\left(0, \sum_{x \in \mathcal{U}, i_T(x) = i} \Pr\left[x \leftarrow A\right] - e^{\varepsilon}\left(\frac{\Pr\left[x \leftarrow A\right]}{f^i} + \left(\Pr\left[x \leftarrow B\right] - \frac{\Pr\left[x \leftarrow A\right]}{f^{i_T(x)}}\right)\right)\right) \\
&= \sum_{i \in \{j_\varepsilon, \ldots, n\}} \max\left(0, \sum_{x \in \mathcal{U}, i_T(x) = i} \Pr\left[x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \leftarrow B\right]\right) \\
&\leq \sum_{i \in \{j_\varepsilon, \ldots, n\}} \sum_{x \in \mathcal{U}, i_T(x) = i} \max\left(0, \Pr\left[x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \leftarrow B\right]\right) \\
&\leq \sum_{x \in \mathcal{U}} \max\left(0, \Pr\left[x \leftarrow A\right] - e^{\varepsilon}\Pr\left[x \leftarrow B\right]\right)
\end{aligned}$$

Hence, we conclude that

$$\delta^{\text{low}} \leq \sum_{x \in \mathcal{U}} \max\left(0, \Pr\left[x \leftarrow A\right] - e^{\varepsilon} \Pr\left[x \leftarrow B\right]\right) = \delta_{\varepsilon}$$

□

Thus, the bounds calculated present a sound over-approximation of the real differential privacy values.

4.5 Implementation

We implemented our algorithm in Python using the NumPy [2] and the SciPy [3] libraries in 740 LoC. The most time consuming part in the computation is the composition. We phrased the composition as a series of inner products and use the NumPy library, which has an efficient implementation of inner products. We added a simple form of parallelization (62 LoC), but expect that a massive parallelization via GPUs should be several orders of magnitudes more efficiency than our current implementation.

Given a bucketing factor as well as a number of buckets $2n + 2$, our implementation constructs bucket distributions from any given histogram / distribution with a limed number of events. For Laplacian noise and Gaussian noise we have implemented special constructors that create bucket distributions for those functions in a more-or less precise fashion.

Given any bucket distribution and a number of rounds r, our implementation then calculates both upper bounds (with error correction) and lower bounds using repeated squaring: we compose the bucket distribution with itself in each round, thus calculating 2^r in a time linear in r (and quadratic in the number of buckets n). Our implementation adaptively decides whether or not to perform "squaring", i.e., to rebase the factor depending on whether the bucket with index ∞ would otherwise grow too much. Empirically, we found that an increase of weight of the ∞ bucket by more than a factor of 2.2 is a good indicator that squaring should be performed. Additionally, we include a parameter that disables squaring as long as the $\mathcal{B}(\infty)$ is below this parameter, which is important for cases where $\mathcal{B}(\infty)$ is initially zero or very small. Finally, we compute an ε, δ-graph by calculating δ as in Definition 14 for every $\varepsilon = f^i$ with $i \in \{0, \ldots, n\}$.

5 Comparison to Kairouz et al.'s composition theorem

Kairouz et al. proved a composition theorem [10] that significantly improves on the standard and advanced composition theorem. This composition theorem [10] provides a composition result where each ε, δ pair after k-fold composition is solely derived from one ε, δ pair of the original pair of distributions. Hence, this composition result does take the entire shape of the distribution into account. In other words, the resulting epsilon and delta bounds are not necessarily tight in the sense of Definition 1.

Recall that we show that our bucketing approach provides an upper and a lower bound and that the distance between these two bounds can be made arbitrarily small by increasing the granularity of the buckets. The bucketing can be seen as an approximation of the two ε, δ graphs[5] of the original pair of distributions A and B. As a consequence, our results show that the two ε, δ graphs of A and B capture all features that are relevant for computing the two ε, δ graphs after k-fold composition (i.e., of A^k and B^k).

We show in this section that Kairouz et al.'s composition theorem seems to be tight for the Laplace mechanisms but not for all mechanisms, such as the Gaussian mechanism or the measured timing-leakage of the CoverUp system [14]. While our approach does not provide significantly tighter bounds for Laplace mechanism, our bucketing significantly improves the privacy bounds on other mechanisms, such as Gaussian mechanism and CoverUp-data. We first describe how we compute these mechanisms and then how we compute the composition theorem. Subsequently, we compare the tightness of the bounds from our bucketing approach to the bounds from Kairouz et al.'s composition theorem in these three scenarios. In the three case studies of this section we consider one-dimensional data, e.g., in responses to statistical queries over sensitive databases or leakage due to suspicious timing delays. However, our approach and our implementation can also deal with higher-dimensional data.

[5] There are two ε, δ graphs since the DP definition is asymmetric.

5.1 Embedding the Laplace mechanism

We analyze the Laplace mechanism, the classical mechanism to achieve DP, by comparing two distributions of Laplace noise with means 0 and 1 respectively. This case corresponds to many applications of the Laplace mechanism for DP, such as counting queries for databases with sensitivity 1. We choose in our case study a Laplace distribution with mean $\mu = 0$ and scale factor $\gamma = 200$, denoted as $\text{LP}(\mu, \gamma)$. As a result, an attacker either makes observations from $\text{LP}(0, 200)$ or from $\text{LP}(1, 200)$ (as the sensitivity is 1). We consider truncated Laplace distributions, since that corresponds closer to real-world applications. If not mentioned otherwise, we truncate at $\mu - 2500$ and $\mu + 2500$.

We want to give strong evidence that both Kairouz et al.'s composition theorem and our bucketing approach are tight for the bounds of the Laplace mechanism. As a consequence, we carefully embed the Laplace mechanism in a way that has a small discretization error. The bucketing method introduced in Definition 9 iterates over all atomic events in the support of the distributions. For modeling the Laplace distribution, or rather, two Laplace distributions A and B, we consider the quotients of the probability mass functions and integrate distribution A over the range of events that fall into each bucket: for $\mathcal{B}(i)$ we integrate over all events E such that $f^i < p_A(E)/p_B(E) \leq f^{i+1}$. This technique can also be applied to other distributions with an infinitely large support, where all areas where B has a probability of zero naturally contribute to the bucket $\mathcal{B}_\infty$.

Recall the probability density function for the Laplace distribution with mean μ and scale parameter γ as $\text{Laplace}(x) := \frac{1}{2\gamma} e^{\frac{-|x-\mu|}{\gamma}}$. For differential privacy we often compare two such distributions with the same scale parameter γ and different medians μ_1 and μ_2, where the means are the real values to which we add Laplace noise with scale parameter γ. We know that without composition, we get $(\varepsilon, 0)$-DP with $\varepsilon = \frac{1}{\gamma}$. Consequently, we can describe the quotient f at each point x as We calculate the quotient $f(x) = \frac{\text{Laplace}_{\mu_1}(x)}{\text{Laplace}_{\mu_2}(x)}$ depending on the relation between the values for x, μ_1 and μ_2:

- $x \leq \min(\mu_1, \mu_2)$: $f(x) = e^{-(\mu_1 - x)\varepsilon} / e^{-(\mu_2 - x)\varepsilon} = e^{(-\mu_1 + x - x + \mu_2)\varepsilon} = e^{(\mu_2 - \mu_1)\varepsilon}$
- $\mu_1 \geq x \geq \mu_2$: $f(x) = e^{-(\mu_1 - x)\varepsilon} / e^{-(x - \mu_2)\varepsilon} = e^{(-\mu_1 + x + x - \mu_2)\varepsilon} = e^{(-\mu_1 - \mu_2 + 2x)\varepsilon}$
- $\mu_1 \leq x \leq \mu_2$: $f(x) = e^{-(x - \mu_1)\varepsilon} / e^{-(\mu_2 - x)\varepsilon} = e^{(-x + \mu_1 + \mu_2 - x)\varepsilon} = e^{(\mu_1 + \mu_2 - 2x)\varepsilon}$
- $x \geq \max(\mu_1, \mu_2)$: $f(x) = e^{-(x - \mu_1)\varepsilon} / e^{-(x - \mu_2)\varepsilon} = e^{(\mu_1 - x + x - \mu_2)\varepsilon} = e^{(\mu_1 - \mu_2)\varepsilon}$

It turns out that for a pair of Laplace distributions the quotient in the region $\min(\mu_1, \mu_2) \leq x \leq \max(\mu_1, \mu_2)$ is either monotonically increasing or monotonically decreasing. For any x smaller than $\min(\mu_1, \mu_2)$, the quotient is stable at $e^{-\varepsilon}$ and for any x larger than $\max(\mu_1, \mu_2)$ the quotient is stable at e^{ε}. Recall that our buckets capture a *range of quotients*: bucket i captures all x such that $f^i < p_A(E)/p_B(E) \leq f^{i+1}$. As a result, each bucket i contains contiguous points and defines an interval on the $x - axis$. For each interval we define the *bucket borders*, i.e., for the bucket with index i, we call the value x with $f(x) = f^{i-1}$ the *left bucket border* $\text{lbb}(i)$ and the value x with $f(x) = f^i$ the *right bucket border* $\text{rbb}(i)$.

For $\mu_1 > \mu_2$, the right bucket border $\text{rbb}(i)$ is the x such that

$$
\begin{aligned}
e^{(2x - \mu_1 - \mu_2)\varepsilon} &= f^i = e^{(i\varepsilon/\text{gr})} =: e^j \\
\Leftrightarrow (2x - \mu_1 - \mu_2)\varepsilon &= j \\
\Leftrightarrow (2x - \mu_1 - \mu_2) &= j/\varepsilon \\
\Leftrightarrow 2x &= \mu_1 + \mu_2 + j/\varepsilon \\
\Leftrightarrow x &= (\mu_1 + \mu_2 + j/\varepsilon)/2 \\
\Leftrightarrow x &= (\mu_1 + \mu_2 + \frac{(i\varepsilon/\text{gr})}{\varepsilon})/2 \\
\Leftrightarrow x &= (\mu_1 + \mu_2 + i/\text{gr})/2 \\
\implies \text{rbb}(i) &= 1/2(\mu_1 + \mu_2 + i/\text{gr}) \\
\implies \text{rbb}(i-1) &= 1/2(\mu_1 + \mu_2 + i/\text{gr} - 1/\text{gr}) \\
&= \text{rbb}(i) - 1/(2\text{gr}) \\
&= \text{lbb}(i)
\end{aligned}
$$

For $\mu_1 < \mu_2$, the right bucket border $\text{rbb}(i)$ is the x such that

$$\begin{aligned}
e^{(-2x+\mu_1+\mu_2)\varepsilon} &= f^i = e^{(i\varepsilon/\text{gr})} =: e^j \\
\Leftrightarrow (-2x+\mu_1+\mu_2)\varepsilon &= j \\
\Leftrightarrow (-2x+\mu_1+\mu_2) &= j/\varepsilon \\
\Leftrightarrow 2x &= \mu_1+\mu_2-j/\varepsilon \\
\Leftrightarrow x &= (\mu_1+\mu_2-j/\varepsilon)/2 \\
\Leftrightarrow x &= (\mu_1+\mu_2-\frac{(i\varepsilon/\text{gr})}{\varepsilon})/2 \\
\Leftrightarrow x &= (\mu_1+\mu_2-i/\text{gr})/2 \\
\implies \text{rbb}(i) &= 1/2(\mu_1+\mu_2-i/\text{gr}) \\
\implies \text{rbb}(i-1) &= 1/2(\mu_1+\mu_2-i/\text{gr}+1/\text{gr}) \\
&= \text{rbb}(i)+1/(2\text{gr}) \\
&= \text{lbb}(i)
\end{aligned}$$

As a result, the bucket i has the value $\int_{\text{lbb}(i)}^{\text{rbb}(i)} \text{Laplace}(\mu_1, 1/\epsilon)$.

We compute the error correction term as $\ell(i) := \int_{\text{lbb}(i)}^{\text{rbb}(i)} \left(B(x) - \frac{A(x)}{f^i}\right)$ and we can directly compute the virtual error from this term.

For the buckets with index $\pm i$ s.t. $f^i = e^\varepsilon$ we integrate over the respective remaining areas $\mathcal{B}(-i) = \int_{-\infty}^{\text{rbb}(-i)} \text{Laplace}(\mu_1, 1/\epsilon)$ and to $\mathcal{B}(i)$ we add $\int_{\text{rbb}(i)}^{\infty} \text{Laplace}(\mu_1, 1/\epsilon)$. As we chose f to fit e^ε the events in these regions exactly have the respective quotient of the bucket and we don't have errors for these integrals. Consequently, the error terms for bucket $\mathcal{B}(-i)$ are zero and the error terms for bucket $\mathcal{B}(i)$ are composed of the error terms for the values x with $\text{lbb}(i) < x < \text{rbb}(i)$.

Truncated Laplace distributions. The truncation of each of either of these functions, causes the quotient of a region to be either 0 or to have 0 in the denominator, which we treat as infinity. The regions are captured by the outer buckets with indexes $-n$ and ∞ respectively.

5.2 Embedding the Gaussian mechanism

The truncated Gaussian mechanism is also an often-used mechanism in privacy-preserving applications. It works similar to the Laplace mechanism insofar as it convolves the input (e.g., a query response) with a Gaussian distribution. In this work, we use a mean $\mu = 0$ and a standard deviation $\sigma = 200\sqrt{2}$ (to achieve the same variance as $\text{LP}(0, 200)$) – denoted as $\text{GS}(\mu, \sigma^2)$ –, and we truncate these distributions at $\mu - 2500$ and $\mu + 2500$, if not mentioned otherwise.

For the truncated Gaussian mechanism, we do not use a precise embedding but rather produce a histogram for each of the two distributions, using SciPy's `scipy.stat.norm` function. Then, we use the normal interface of our bucketing implementation that parses a pair of histograms and produces a bucketlist vector, a real error vector, and a virtual error vector. We accept that this implementation may produce significant discretization artifacts that, however, should be both small w.r.t. the values concerned and should not lead to a significantly different shape of the distributions under composition.

5.3 Embedding CoverUp's data

We illustrate the expressivity of our approach by applying it to measured data from the CoverUp paper, where the input signal is a distribution of response delays to which Gaussian noise has been added. Hence, the original signal (the response delays) are convolved with a Gaussian distribution. CoverUp uses the Gaussian distribution $\text{GS}(0, 200^2)$ that is truncated at -1000ms and 1000ms. This use case shows that our approach can be applied to the analysis of complex distributions as easily as to randomized responses to database queries with a fixed sensitivity.

Classical anonymous communication networks (ACN) have the goal of hiding the IP address of the sender and the recipient of a communication. Such ACNs do however not hide the participation time, i.e., whether, when, and for how long a party uses an ACN. This participation time can be used for long-term attacks (e.g., intersection attacks) and can raise suspicion national state-level adversaries. Sommer et al. [14] propose a system, called CoverUp, that has the goal of hiding this participation time leakage. CoverUp assumes a collaborating popular web service with a significant amount of regular visitors. This webpage would be incorporated into the usage of an ACN and trigger all its visitors to produce cover traffic. This web page would serve an iFrame that loads content from a trusted server, which in turn would serve a piece of JavaScript code that executes a dummy client for the ACN on the visitors browser. ACN users would act as a normal visitor, receive the JS code, but additionally have a dedicated CoverUp browser extension installed. The browser extension would enable a communication channel to an external application by replacing the dummy messages from the dummy client with actual messages from an external application and by forwarding all messages from the network to the external application. For CoverUp to properly hide the participation time ACN users (called *voluntary* participants) and normal website visitors (called *involuntary* participants) have to be indistinguishable. While both execute the same piece of JS code, the voluntary participants perform additional computations. As a consequence, the response time of the voluntary participants differs by a few milliseconds from the response time of the involuntary participants. CoverUp remedies this timing leakage by adding random delays in the JS code, i.e., for voluntary and involuntary participants.

The CoverUp paper presents an analysis of this timing leakage (after adding the noise) and aims for a high degree of privacy after more than 250k observations. The CoverUp authors experimentally measured the timing delays of voluntary and involuntary participants in the lab and produced histograms of these timing delays. These histograms are used as a model for the timing delays of voluntary and involuntary participants to assess the timing leakage of CoverUp. We apply our algorithm to these histograms of timing delays, to illustrate that and how well our approach works on measured data. We use data from the CoverUp project, which is openly available online.[6]

In this comparison, we only consider those measured delays on a Linux system that are observable after the webpage has been loaded, called the "periodic" measurements in the CoverUp paper.

5.4 Computing the Kairouz et al.'s composition theorem

We directly implement Kairouz et al.'s composition theorem.

Theorem 3 ([10]). *For any $\varepsilon \geq 0$ and $\delta \in [0,1]$, the class of (ε, δ)-differentially private mechanisms satisfies*

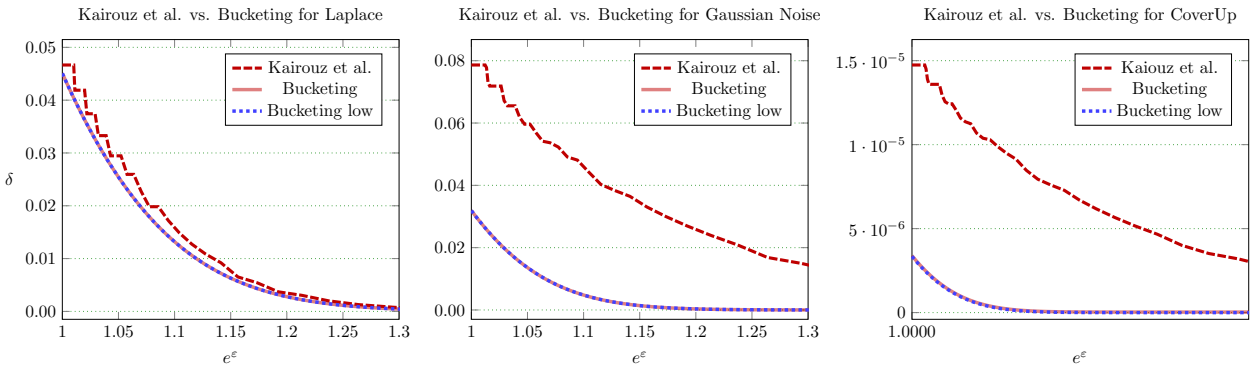
$$(\varepsilon', \delta')\text{-differential privacy}$$

under k-fold composition, for all $i \in \{0, \ldots, \lfloor k/2 \rfloor\}$ where $\varepsilon' = (k-2i)\varepsilon$ and $\delta' = 1 - (1-\delta)^k(1-\delta_i)$

$$\delta_i = \frac{\sum_{\ell=0}^{i-1} \binom{k}{\ell} \left(e^{(k-\ell)\varepsilon} - e^{(k-2i+\ell)\varepsilon}\right)}{(1+e^{\varepsilon})^k}$$

We compute for a given k the composition by looking up for a fine-grid of ε values the corresponding δ value of the original pair of distributions and then computing and storing all (ε', δ') pairs according to the theorem above, i.e., for all $i \in \{0, \ldots, \lfloor k/2 \rfloor\}$. From these stored (ε', δ') pairs, we remove all pairs for which we have stored lower $(\varepsilon'', \delta'')$ pairs, i.e., pairs such that $\varepsilon'' \leq \varepsilon'$ and $\delta'' \leq \delta'$. We output the remaining list of (ε', δ') pairs, which form a monotonically decreasing (ε, δ)-graph. Due to our direct implementation of δ_i, we can only evaluate the composition theorem up to $k = 512$ before the intermediate computation results (in particular, the $e^{O(k)}$-terms) become too large.

In our computation, the granularity of the grid of ε values of the original pair of distributions naturally leads to an imprecision. We use a fine grid of $e^{\varepsilon} \in \{(1+10^{-14})^{1.1^j} \mid j \in \{0, \ldots, n\}\}$, where we choose n as a point where the (ε, δ) after k-fold composition becomes stationary. While we concede that it might be possible to obtain slightly lower bound from the composition theorem, we are confident that due to this fine grid the resulting graphs for Kairouz et al's composition theorem that we compute are representative.



(a) Laplace mechanism (b) Gaussian mechanism (c) CoverUp

Figure 6: The ε, δ graphs computed with Kairouz et al.'s composition theorem and with our bucketing approach after $k = 512$ compositions for the Laplace mechanism, the Gaussian mechanism, and the CoverUp data. The y-axis depicts the δ-values and the x-axis the e^ε values. The variance of the Gaussian mechanism and the Laplace mechanism is $80,000$, the sensitivity is 1 (the centers at $\mu_1 = 0$ and $\mu_2 = 1$) respectively, and in both mechanisms truncation happens at -2500 and $+2500$ from the respective μ_i.

5.5 Comparing evaluations

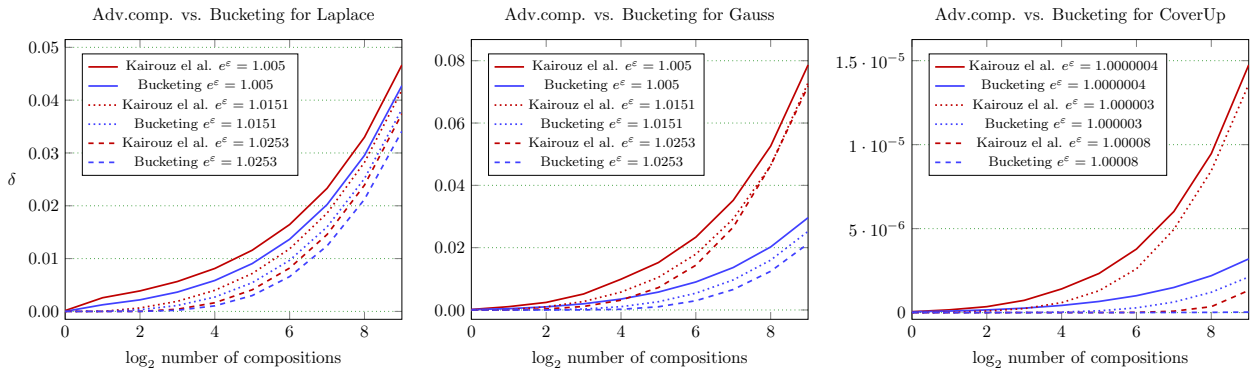
We are finally in a position to evaluate how our bucketing approach compares against Kairouz et al.'s composition theorem. Figure 6a shows that our upper and lower bounds coincide, i.e., our results are tight. Also, Kairouz et al.'s composition theorem is tight with respect to a pair of Laplace distributions (i.e., the Laplace mechanism). Figure 6b shows that for the Gaussian mechanism that composition theorem is already after 512 compositions not very tight. Figure 6c shows that for the CoverUp-data our bucketing approach is tight, while there is a large gap to the bounds from Kairouz et al's composition theorem.

Figure 7 compares for fixed epsilon values the evolution of the delta bounds from Kairouz et al.'s composition theorem and from our approach. This comparison again uses the Laplace mechanism, the Gaussian mechanism and the CoverUp data.

6 Comparison of the Gaussian and the Laplace mechanism

As we have seen in Section 5.5, Kairouz et al.'s composition theorem is fairly tight for the Laplace mechanism but not for the Gaussian mechanism. Figure 8 (upper two graphs) compares a truncated Laplace and a truncated Gaussian mechanism and find that for the same variance the Gaussian mechanism provides a significantly higher degree of privacy. For a fixed variance of $80,000$, a sensitivity of 1 ($mu_1 = 0$ and $\mu_2 = 2$), and a truncation at -2500 and 2500 for μ_1 (and -2499 and 2501 for μ_2), the upper left graph in Figure 8 depicts how, for different but fixed epsilon values, the delta increases over the course of 512 evaluations. The graph clearly shows that in the course of 512 compositions, the reduced leakage of the Gaussian mechanism becomes visible. The upper right graph in Figure 8 shows the full epsilon-delta graphs of a Gaussian and a Laplace mechanism after 512 compositions, where the two mechanisms use noise that has the same variance $(80,000)$. In particular, the delta-value where the (ε, δ) graph levels out is 4 orders of magnitude lower for Gaussian noise than it is for Laplace noise, since the Gaussian distribution falls much steeper than Laplace distribution. This difference of the Gaussian and the Laplace mechanisms becomes even more pronounced in our analysis and improvement of the Vuvuzela protocol in Section 7. The analysis of Vuvuzela also illustrates that the steepness of the Gaussian distribution enables a much tighter truncation, i.e., the distribution can be truncated much earlier than a Laplace distribution without sacrificing privacy. This tighter truncation, in turn, leads to a smaller range of noise that is required to achieve the same privacy goals as with Laplace noise.

[6] Available under `http://coverup.tech`.



(a) Laplace (b) Gauss (c) CoverUp

Figure 7: The ε, δ graphs computed with Kairouz et al.'s composition theorem and with our bucketing approach after $k = 512$ compositions, on the left hand side applied to a pair of Laplace distributions with a scale factor of $\gamma = 200$ and $\mu_1 = 0$ and $\mu_2 = 1$ (together with the lower bound) and on the right hand side applied to a pair of histograms of timing-leakage-data measured from a real system (without the lower bound). The y-axis depicts the δ-values and the x-axis the e^ε values.

Additionally, we found surprising evidence that the epsilon-delta graph of the Laplace mechanism converges toward the epsilon-delta graph of the Gaussian mechanism, as long as the Gaussian mechanism has half the variance of the Laplace mechanism. For the same sensitivity, and truncations as above, the lower two graphs in Figure 8 illustrate that after 512 compositions these two graphs converge toward each other. The lower left graph in Figure 8 depicts how, for different but fixed epsilon values, the delta increases over the course of 512 evaluations. The graph clearly shows how in the course of 512 compositions, the delta values of the Laplace mechanism converge toward the delta values of the Gaussian mechanism. The lower right graph in Figure 8 shows the full epsilon-delta graphs of a Gaussian and a Laplace mechanism after 512 compositions, where the Laplace mechanism has twice the variance $(80,000)$ of the Gaussian mechanism $(40,000)$. This figure shows how close the two epsilon-delta graphs are and that they almost only differ due to their different y-values at the point where they have been truncated. This difference, however, is crucial. As explained above, it is caused by the steepness of the Gaussian distribution and enables a much tighter truncation, which in turn can lead to significantly less noise overhead, as we illustrate in our analysis of Vuvuzela. We leave it for future work to investigate this connection further.

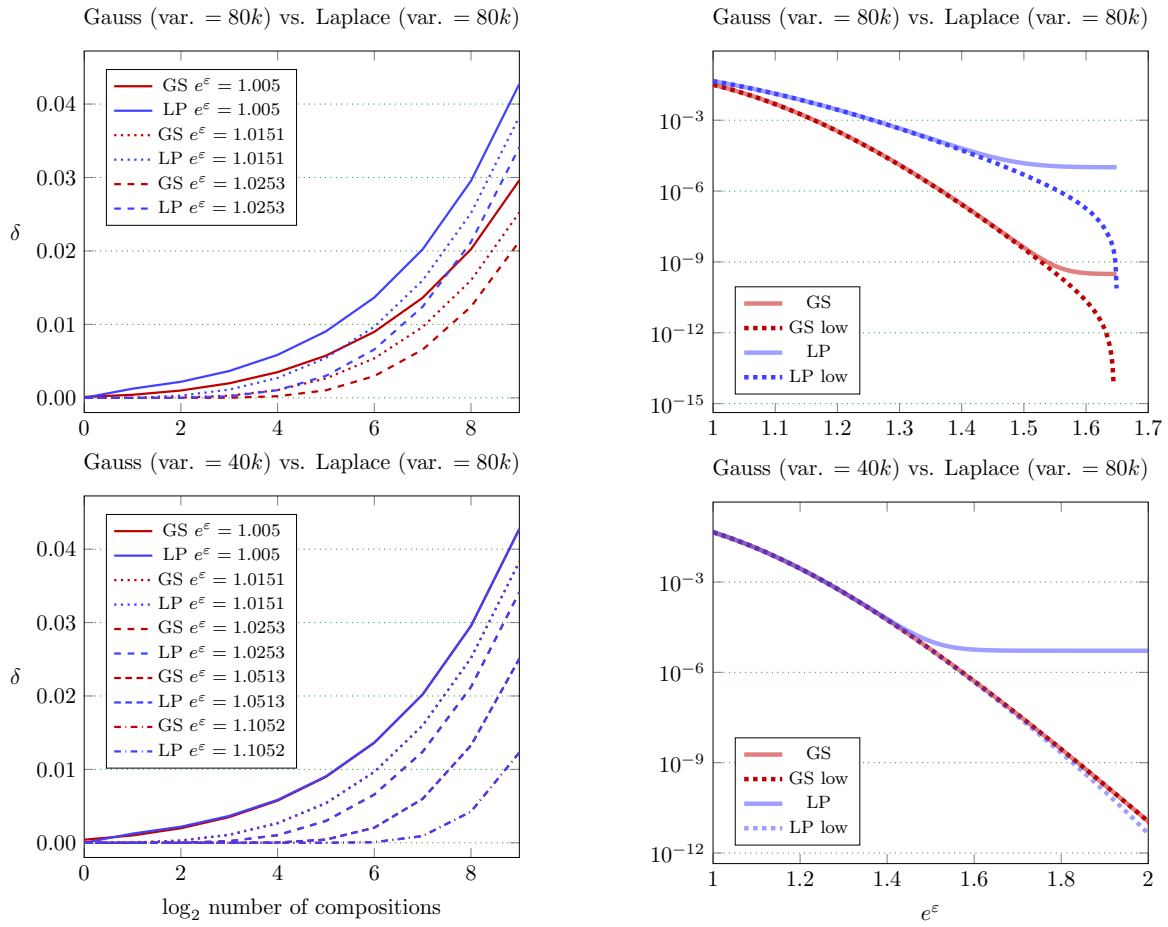
7 Application to Vuvuzela

In this section, we show how aiming for tight bounds in a privacy analysis can significantly improve the bandwidth overhead of a protocol. We use the Vuvuzela [15] protocol as an example, which is an anonymous communication system tailored towards messengers. Vuvuzela uses Laplace noise to achieve strong privacy properties. Using the insights from Section 6, we go one step further and propose to change the shape of the noise distribution from Laplace noise to Gaussian noise. With our bucketing approach, we show that already 5 to 10 times less noise[7] suffices to achieve the same strong privacy properties.

We acknowledge that for the analysis of the Laplace noise Kairouz et al.'s composition result would already yield significantly better results than Vuvuzela's original analysis. For the analysis of the Gaussian noise, however, our comparison from Section 6 indicates that their result would not have provided tight guarantees.

We refer to the original Vuvuzela paper for a full presentation and restrict our presentation to the bare bones that are needed to understand the noise messages that Vuvuzela uses to achieve strong privacy properties.

[7]The more observations are estimated, the higher the error of the advanced composition result, which is used in the original analysis from the Vuvuzela paper; hence, in those cases the tightness of our bounds leads to a more significant improvement.



(a) Growth of δ over the number of compositions (y-axis) for fixed epsilon values (different line-styles) for a growing number of compositions (x-axis in $\log_2$-scale).

(b) The ε, δ graphs (upper and lower bounds) after $k = 512$ compositions applied to a Gaussian and a Laplace mechanism with δ on the y-axis and e^ε on the x-axis.

Figure 8: Truncated Gaussian mechanisms (red) vs. truncated Laplace mechanism (blue) both with sensitivity $= 1$. For both mechanism truncation is at $\mu_i - 2500$ and $\mu_i + 2500$ ($\mu_1 = 0$ and $\mu_2 = 1$). At twice the variance the Laplace mechanism converges towards the Gaussian mechanism, so much that the blue lines almost completely cover the red lines.

We stress that our work contributes to improving the epsilon-delta bounds and thus to improve a given privacy analysis. This work is not meant to help in finding a suitable attacker model, a suitable definition or accurate usage profiles. Hence, we stick to Vuvuzela's privacy analysis, as it was presented in the original paper.

7.1 Protocol overview

Vuvuzela clients communicate by deposing their encrypted messages in virtual locations in the one of the mixes (the locations are called *dead drops*). For agreeing on such a dead drops, Vuvuzela deploys a dialing protocol where the dialer sends the ID of a dead drop to dedicated invitation dead drops. This ID is encrypted with the peer's public key with an encryption schemes that is designed to hide the recipient's identity. On the dialer's side directly the conversation protocol is started where the client regularly retrieves the chat messages from and deposits chat messages to the dead drop from the invitation. If the recipient receives and accepts the invitation, the recipient also starts the conversation protocol.

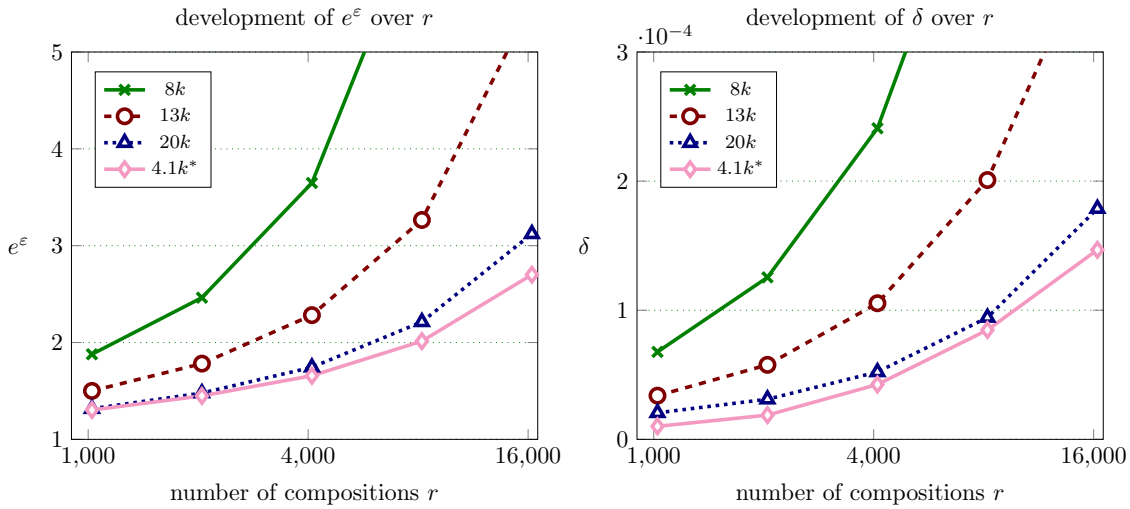
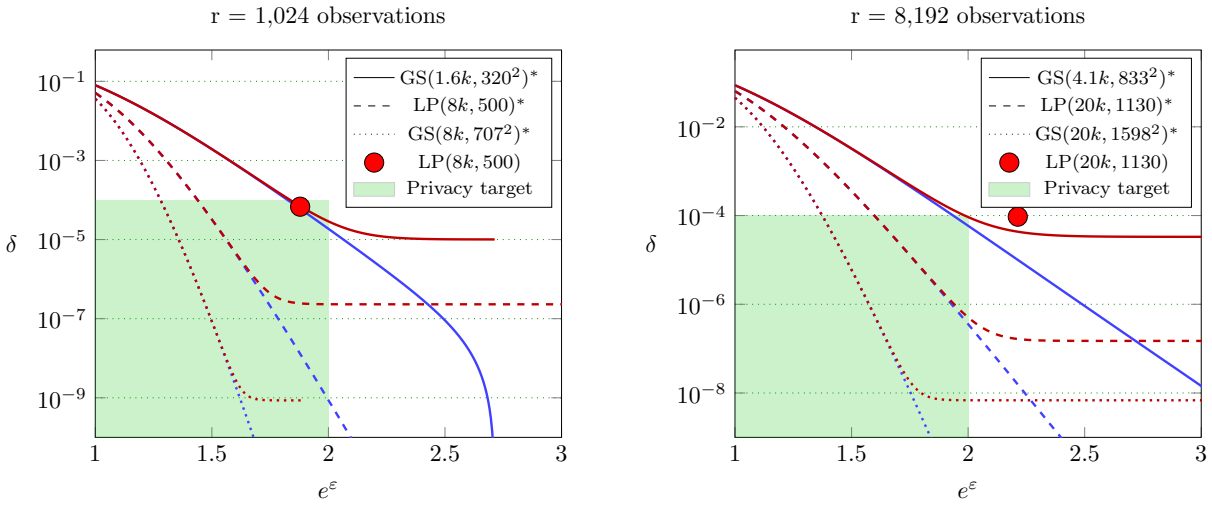


Figure 9: The privacy bounds for Vuvuzela's dialing protocol. The left graph shows the e^{ε}-values on the y-axis and the number of observations r on the x-axis (i.e., r-fold composition) and the right graph shows the corresponding δ-values on the y-axis. The solid green ($\mu = 8k, \gamma = 500$), the dashed red ($\mu = 13k, \gamma = 770k$), and the dotted blue line ($\mu = 20k, \gamma = 1130$) are from the original Vuvuzela paper, and the solid magenta line (Gaussian noise, $\mu = 4.1k^*, \sigma = 320$) is computed with this work's technique.

Privacy analysis. Vuvuzela assumes a global network-level attacker that is additionally able to compromise some mixes. To achieve strong resistance against compromised servers, each path in Vuvuzela traverses all nodes. To counter traffic correlation attacks, Vuvuzela clients produce dummy trafic at a constant rate. The Vuvuzela paper argues that the only remaining source of leakage is the patterns of registering invitations and patterns of access requests to these dead drops: single requests to dead drops, corresponding to dummy messages or messages before the peer accepted the conversation, and pairs of requests to the same dead drop, corresponding to an active conversation.

Privacy-enhancing measures. Vuvuzela reduces the information that an attacker can learn by triggering each mix to produce cover stories for potentially communicating parties. For the dialing protocol, the mixes produce cover stories (i) by sending dummy invitation registrations and invitation requests to the dedicated invitation dead drops. The number of these dummy registrations and dummy requests is in each round drawn from the truncated Laplace distribution $\lceil \max(0, \text{Laplace}(\gamma_d, \mu_d)) \rceil$ for some system parameters γ_d and μ_d. For the conversation protocol, the mixes produce cover stories (ii) for idle parties, by sending pairs of dummy access requests to uniform-randomly chosen dead drops, and (iii) for (bi-directionally) communicating parties, by sending (single) dummy access requests to uniform-randomly chosen dead drops. The number of (single) dummy access requests (ii) is in each round drawn from the truncated Laplace distribution $\lceil \max(0, \text{Laplace}(\gamma_c, \mu_c)) \rceil$ for a system parameters γ_c and μ_c, and the number of pairs of dummy access requests (iii) is in each round drawn from the truncated Laplace distribution $\lceil \max(0, \text{Laplace}(\gamma_c/2, \mu_c/2)) \rceil$. The system parameters $\gamma_d, \gamma_c, \mu_d, \mu_c$ determine how much noise-overhead the protocol produces and how much privacy it will offer.

Privacy-impact of the dummy requests. The goal of the these dummy requests and invitations is to produce a cover stories for dialing parties (i), for idle parties (ii), and for conversing (iii). The Vuvuzela paper separately conducts a privacy analysis for the dialing protocol ((i)) and the conversation protocol ((ii) and (iii) combined). For the dialing protocol, the paper concludes that it suffices to bound the r-fold (ϵ, δ) differential privacy of $\max(0, \text{Laplace}(\gamma_d, \mu_d))$ and $\max(0, \text{Laplace}(\gamma_d, \mu_d + 2))$, i.e., the (ϵ, δ) differential privacy of the product distributions $\max(0, \text{Laplace}(\gamma_d, \mu_d))^r$ and $\max(0, \text{Laplace}(\gamma_d, \mu_d + 2))^r$. The parameter r indicates the number of rounds at which that the attacker conducts an observation. For the conversation protocol, the paper concludes that it suffices to estimate the r-fold (ϵ, δ) differential privacy of $\max(0, \text{Laplace}(\gamma_c, \mu_c)) +$



(a) After $r = 1,024$ observations with Gaussian noise with $\mu = 1.6k$ and $\sigma = 320$ (solid), Laplace noise $\mu = 8k, \gamma = 500$ (dashed), and Gaussian noise with $\mu = 8k$ and $\sigma = 707$ (dotted), and the red dot represents the ε, δ combination for $\mu = 8k, \gamma = 500$ from the original Vuvuzela paper.

(b) After $r = 8,192$ observations with Gaussian noise with $\mu = 4.1k$ and $\sigma = 833$ (solid), Laplace noise $\mu = 20k, \gamma = 1130$ (dashed), and Gaussian noise with $\mu = 20k$ and $\sigma = 1598$ (dotted), and the red dot represents the ε, δ combination for $\mu = 20k, \gamma = 1130$ from the original Vuvuzela paper.

Figure 10: The (ε, δ) graphs (y-axis and x axis, respectively, y-axis in $\log_{10}$-scale) from our method in comparison with the bound from the original Vuvuzela paper (for the dialing protocol). The figure depicts upper (red) and a lower bounds (blue) and Vuvuzela's privacy target (green, $\delta \leq 10^{-4}, e^{\varepsilon} \leq 2$).

$\max(0, \text{Laplace}(\gamma_c/2, \mu_c/2))$ and $\max(0, \text{Laplace}(\gamma_c, \mu_c+2)) + \max(0, \text{Laplace}(\gamma_c/2, \mu_c/2+1))$. The Vuvuzela paper uses the advanced composition theorem for differential privacy [7] to bound ϵ and δ. The paper analyzes for the conversation protocol three system parameters: $\mu = 150k, \gamma = 7.5k$, $\mu = 300k, \gamma = 13.8k$, and $\mu = 450k, \gamma = 20k$. We show that the resulting bounds can be significantly improved and we indicate all new bounds with a "$*$" sign in the respective figures.

7.2 Tighter privacy analysis for the dialing protocol

We apply our method to estimate tighter ε and δ bounds for Vuvuzela, and to reduce the recommended noise. Recall that we observed in Section 6 that Gaussian noise for the same variance behaves better under composition than Laplacian noise. This section studies how much our tighter bounds enable us to reduces the noise in the case that Gaussian noise is used or that Laplace noise is used, and this section studies how much the originally recommended amount of noie improves the degree of privacy, in case Gaussian noise is used or Laplace noise is used. We stress that while in the case of Vuvuzela there is no utility function that we have to preserve other than to minimize the bandwidth overhead, our approach is also suited for applications where a utility function has to be preserved. In those cases, we would probably reduce the variance to an appropriate level and then compute tight bounds.

For the dialing protocol, we show that with Gaussian noise the noise rate can be reduced by a factor of almost 5 while still meeting the privacy requirements, and for the conversation protocol the noise rate can be reduced by a factor of 10 while still meeting the privacy requirements. With Laplace noise the noise rate can be reduced by a factor of 2 and for the conversation protocol by a factor of 4. As we only present the Laplace noise for comparison, we placed the graphs for the Laplace noise (Figure 13 and 14) at the appendix, for the sake of brevity. As the conversation protocol produces more observations (i.e., more compositions) and the untightness of the bounds that the original Vuvuzela paper used amplifies more heavily for a high the number of observations, the tightness of our bounds is more pronounced for the conversation protocol.

For comparability, we depict in Figure 9 the original graphs from the Vuvuzela analysis, which show the epsilon graph and the delta graph with increasing r, respectively, for the dialing protocol and estimated with

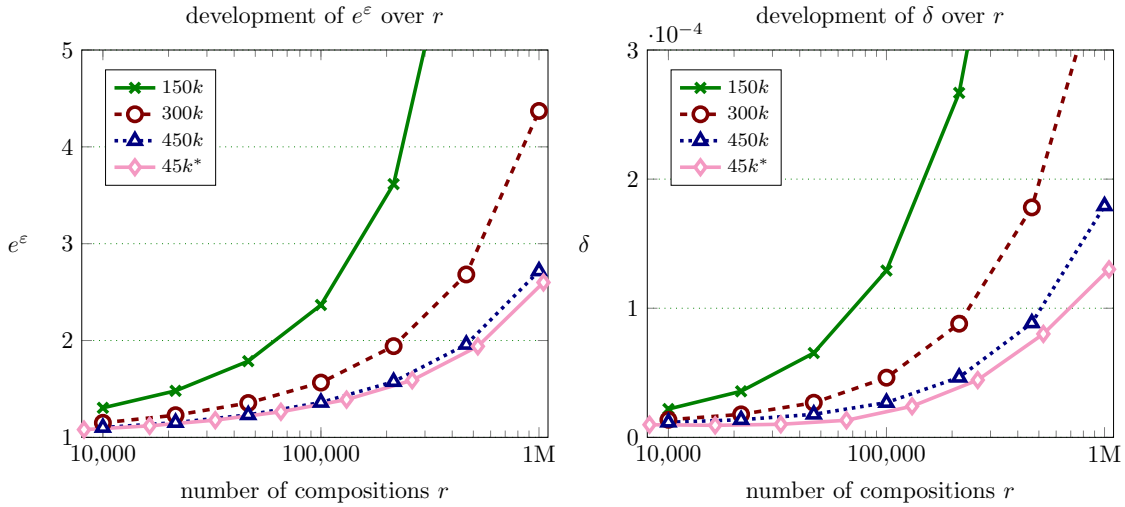


Figure 11: The privacy bounds for Vuvuzela's conversion protocol. The left graph shows the e^ε-values on the y-axis and the number of observations r on the x-axis (i.e., r-fold composition) and the right graph shows the corresponding δ-values on the y-axis. The solid green line LP$(150k, 7.5k)$, the dashed red line LP$(\mu = 300k, \gamma = 13.8k)$, and the dotted blue line LP$(450k, 20k)$ are from the original Vuvuzela paper, and the solid magenta line GS$(45k^*, 7.5k^2)$ is Gaussian noise for which the (ϵ, δ) pairs have been computed with this work's technique.

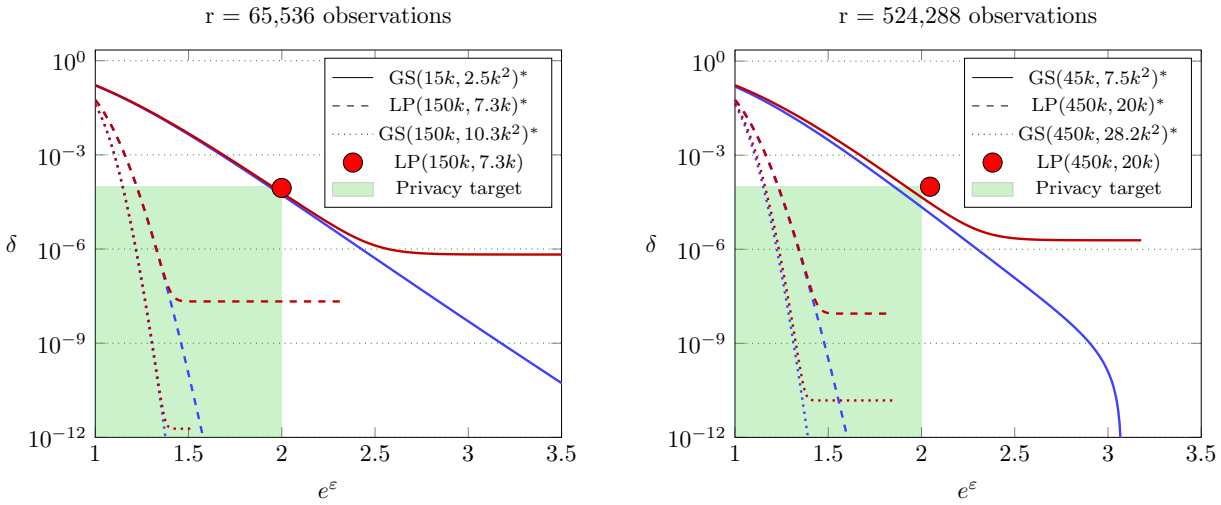
the advanced composition result. We extend those Figures with the lowest, magenta graphs (marked with a $*$) that show the performance of our proposed Gaussian noise that uses nearly 5 times less noise and is computed with our bucketing approach. As our method computes not only one ε, δ pair for each number of observations r but an entire ε, δ graph, we chose representative ϵ values that are close to (and even below) the epsilon values for the highest noise configuration LP$(20k, 1130)$ from the original Vuvuzela paper. The figure shows that our bounds with the reduced noise and with using Gaussian noise GS$(4.1k, 833^2)$ are below the previous bounds for the highest noise configuration LP$(20k, 1130)$, proving that a noise reduction of nearly a factor of 5 still yields for the dialing protocol to achieve the privacy requirements of $e^\varepsilon \leq 2$ and $\delta \leq 10^{-4}$.

Next, we illustrate that our method computes bounds that are several orders of magnitude better than Vuvuzela's original bounds. For $r = 8,192$ observations, Figure 10b illustrates that using the highest noise configuration with Laplace noise LP$(20k, 1130)$ results in a privacy bound that is almost 3 orders of magnitude lower, in terms of the delta, and with Gaussian noise GS$(20k, 1598^2)$ more than 4 orders of magnitude. The figure depicts the ε, δ graphs computed by our approach for the highest noise configuration LP$(20k, 1130)$, for the corresponding Gaussian noise GS$(20k, 1598^2)$, for the configuration that we propose GS$(4.1k, 833^2)$), and compares it against Vuvuzela's previous bounds LP$(20k, 1130)$. We additionally depict the respective lower bounds, which show that our bounds are quite tight in the sense that there is not much room for improvement. Moreover, due to the more comprehensive view that a full ε, δ graph provides, we can see that the the highest noise configuration with Gaussian noise GS$(20k, 1598^2)$ even achieves the privacy requirements ($\delta \leq 10^{-4}$) for less than $e^\varepsilon = 1.5$ after $8,192$ observations.[8]

We would like to stress that the lower bounds show that our result is tight up to $\delta \geq 10^{-4}$ for GS$(4.1k, 833^2)$, $\delta \geq 10^{-6}$ for LP$(20k, 1130)$, and GS$(20k, 1598^2)$ for $\delta \geq 10^{-8}$. This tightness is solely a scalability issue and ultimately only depends on the number (and hence granularity) of the buckets. A more optimized implementation (e.g., based on GPUs) would be able to significantly increase the number of buckets, thus achieving even tighter upper and lower bounds.

For completeness, we also show in Figure 10a the ε, δ graphs for the dialing protocol for low r: $r = 1024$ and the recommended parameters $\mu = 8k, \gamma = 500$. Here, we can see that our bound is 2 orders of magnitude lower than Vuvuzela's previous bounds for the noise level. The figure also shows that reducing the noise by

[8]Recall that the variance of GS$(\mu, (\sqrt{2}x)^2) = 2x^2$ equals the variance of LP$(\mu, x) = 2x^2$.



(a) After $r = 65,536$ observations with Gaussian noise with $\mu = 15k$ and $\sigma = 2.5k$ (solid), Laplace noise $\mu = 150k, \gamma = 7.3k$ (dashed), and Gaussian noise with $\mu = 150k$ and $\sigma = 10.3k$ (dotted), and the red dot represents the ε, δ combination for $\mu = 150k, \gamma = 7.3k$ from the original Vuvuzela paper.

(b) After $r = 524,288$ observations with Gaussian noise with $\mu = 45k$ and $\sigma = 7.5k$ (solid), Laplace noise $\mu = 450k, \gamma = 20k$ (dashed), and Gaussian noise with $\mu = 450k$ and $\sigma = 28.2k$ (dotted), and the red dot represents the ε, δ combination for $\mu = 450k, \gamma = 20k$ from the original Vuvuzela paper.

Figure 12: The (ε, δ) graphs (y-axis and x axis, respectively, y-axis in $\log_{10}$-scale) from our method in comparison with the bound from the original Vuvuzela paper (for the dialing protocol). The figure depicts upper (red) and a lower bounds (blue) and Vuvuzela's privacy target (green, $\delta \leq 10^{-4}, e^\varepsilon \leq 2$).

a factor of 5, i.e., GS$(1.6k, 320)$, still achieves the privacy requirements ($e^\varepsilon \leq 2$ and $\delta \leq 10^{-4}$).

As a comparison, using Laplace noise only enables a noise reduction of a factor of 2, as shown in Figure 13 in the appendix. Interestingly, the reduced Laplace noise achieves the same privacy bounds as the reduced Gaussian noise if the Laplace noise has twice the variance as the Gaussian noise (i.e., $\gamma = \sigma$) but a 2.5 times wider range, as indicated in Section 6. This shows what a significant effect the steepness of the Gaussian noise can have in practice.

On a technical note, for these truncated Laplace-distributions the (ϵ, δ) bounds are the same no matter whether the distributions are swapped or not; hence, we only compute the bounds for one order.

7.3 Tighter privacy analysis for the conversation protocol

Figure 11 depicts the epsilon graph and the delta graph with increasing r, respectively, for the conversation protocol. We compare Gaussian noise GS$(45k, 28.8^2)k$ with the previous bounds for the recommended noise configurations. We again chose representative ϵ values that are close and even below to the epsilon values for the highest noise configuration LP$(450k, \gamma = 20k)$ from the original Vuvuzela paper and the corresponding Gaussian noise GS$(450k, 28.8k^2)$ for the highest noise configuration. The figure shows that our bounds GS$(45k, 2.5k^2)$ are below the previous bounds for the highest noise configuration LP$(450k, 7.3k)$, proving that a noise reduction by a factor of 10 is sufficient for the conversation protocol to achieve the privacy requirements of $e^\varepsilon \leq 2$ and $\delta \leq 10^{-4}$.

The gap between our bounds and the previous Vuvuzela-bounds is even more pronounced in the analysis of the conversation protocol. For $r = 524,288$ observations, Figure 12b shows that using the highest noise configuration LP$(450k, 20k)$ results in privacy bounds that are almost 4 orders of magnitude lower, in terms of the delta, and for the corresponding Gaussian noise GS$(450k, 28.8k^2)$ more than 6 orders of magnitude. The figure depicts the ε, δ graph for the highest noise recommended configuration LP$(450k, 20k)$), the corresponding Gaussian noise GS$(450k, 28.8k)$, the drastically reduced lowest noise configuration GS$(45k, 7.5k^2)$), and compares these against Vuvuzela's previous bounds. Also, Figure 12b shows the corresponding lower bounds. We can see that our bounds for the reduced noise configuration GS$(45k, 7.5k)$ are tight up to

$\delta \geq 10^{-5}$, for LP$(450k, 20k)$ up to $\delta \geq 10^{-8}$, and for GS$(450k, 28.8k)$ up to $\delta \approx 10^{-10}$. Moreover, we can see that the highest recommended noise configuration with Gaussian noise GS$(450k, 28.8k^2)$ even meets and exceeds the privacy requirements ($e^\varepsilon = 1.25, \delta = 10^{-4}$ or $e^\varepsilon = 1.45, \delta = 10^{-10}$) for $r = 524,288$ observations.

For completeness, we also show in Figure 12a the ε, δ graphs for the conversation protocol for low $r = 65,536$ with noise configurations LP$(150k, 7.3k)$, GS$(150k, 10.3k^2)$, and the factor-10-reduced configuration GS$(15k, 2.5k^2)$. Here, we can also see the tightness of our bound: for LP$(150k, 7.3k)$ up to $\delta \geq 10^{-7}$, for GS$(150k, 10.3k^2)$ up to $\delta \geq 10^{-11}$, and for GS$(15k, 2.5k^2)$ up to $\delta \geq 10^{-6}$. Moreover, we can see that the highest noise configuration GS$(150k, 10.3k^2)$ is more than 7 orders of magnitude lower than Vuvuzela's previous bounds for the same noise level. Moreover, we can see that the highest recommended noise configuration with Gaussian noise GS$(450k, 28.8k^2)$ even meets and exceeds the privacy requirements ($e^\varepsilon = 1.25, \delta = 10^{-4}$ or $e^\varepsilon = 1.4, \delta = 10^{-11}$) for $r = 65,536$ observations.

As a comparison, using Laplace noise only enables a noise reduction of a factor of 4, as shown in Figure 14 in the appendix. Also here, we can observe that the Laplace noise has twice the variance of the Gaussian noise and has a 2.5 times wider range, illustrating the advantages of Gaussian noise in practice.

8 Conclusion and future work

In this paper we have presented *bucketing*, a sound numerical approach for computing upper and lower bounds for differential privacy after k-fold composition. Our approach is based on concrete distributions, but can be applied in a variety of cases, which can include adaptive composition, evolving sequences of distributions and static distributions. All compositions, as well as our reshaping operation of *squaring* the bucket factor have been shown sound and (empirically) tight in many cases.

We applied our bucketing approach to the anonymity network Vuvuzela where we computed bounds for more than half a million compositions, deriving significantly better results than their previous analysis and we found that by exchanging the Laplace noise with Gaussian noise, even better results can be achieved. We also compared our approach to the Kairouz et al.'s composition theorem and found that their theorem provides reasonably tight bounds for the Laplace mechanism but not for other distributions, such as the Gaussian mechanism or for a pair of histograms of timing-leakage measurements from the CoverUp system. We also observed that Gaussian mechanism behaves much better under a high number of compositions than a Laplace mechanism wiht the same variance, and we found evidence that the (ε, δ) graph of the Gaussian mechanism with half the variance of the Laplace mechanism converges to the (ε, δ) graph of the Laplace mechanism.

We encourage the application of our bucketing to other DP mechanisms, such as to the optimal DP mechanisms [8, 11] (e.g., comparing their composition behavior to the Gaussian mechanism) and to privacy-preserving ML methods [1], as well as to improve existing privacy analyses. We consider exploring the relationship (ε, δ)-DP of the Gaussian mechanism and the Laplace mechanism, as well as analyses probing the development of DP provided by other noise distributions under composition an interesting direction for future work.

9 Acknowledgement

This work has been partially supported by the European Commission through H2020-DS-2014-653497 PANORAMIX, the EPSRC Grant EP/M013-286/1, and the Zurich Information Security Center (ZISC).

References

[1] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pages 308–318. ACM, 2016.

[2] N. developers. Nmpy.org: Scientific computing with python. Accessed in August 2017, available at `http://www.numpy.org`.

[3] S. developers. Scipy.org: Scientific computing tools for python. Accessed in August 2017, available at `https://www.scipy.org`.

[4] C. Dwork. Differential privacy. In *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II*, pages 1–12, 2006.

[5] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In *Eurocrypt*, volume 4004, pages 486–503. Springer, 2006.

[6] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(34):211–407, 2014.

[7] C. Dwork, G. N. Rothblum, and S. Vadhan. Boosting and differential privacy. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 51–60. IEEE, 2010.

[8] Q. Geng and P. Viswanath. The optimal mechanism in differential privacy. In *2014 IEEE International Symposium on Information Theory (ISIT)*, pages 2371–2375. IEEE, 2014.

[9] M. Hardt and G. N. Rothblum. A Multiplicative Weights Mechanism for Privacy-Preserving Data Analysis. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 61–70, 2010.

[10] P. Kairouz, S. Oh, and P. Viswanath. The composition theorem for differential privacy. *IEEE Transactions on Information Theory*, 63(6):4037–4049, 2017.

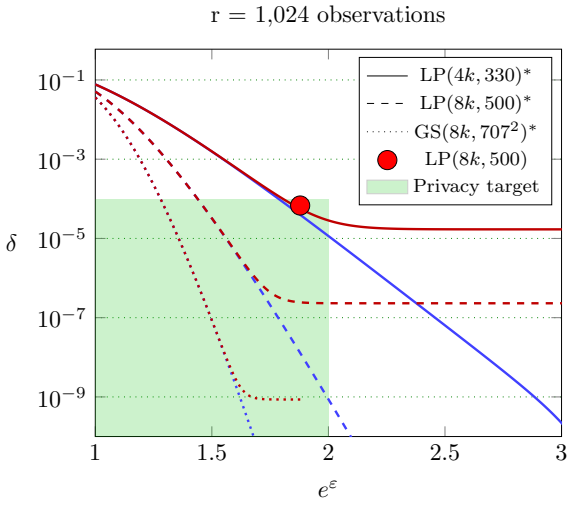
[11] K. Kalantari, L. Sankar, and A. D. Sarwate. Optimal differential privacy mechanisms under hamming distortion for structured source classes. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 2069–2073. IEEE, 2016.

[12] C. Liu, S. Chakraborty, and P. Mittal. Dependence Makes You Vulnberable: Differential Privacy Under Dependent Tuples. In *NDSS*, 2016.

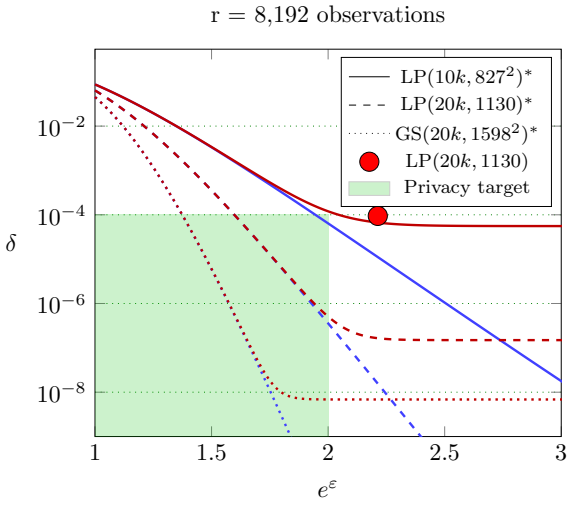
[13] I. Mironov, O. Pandey, O. Reingold, and S. Vadhan. Computational differential privacy. In *Advances in Cryptology-CRYPTO 2009*, pages 126–142. Springer, 2009.

[14] D. Sommer, A. Dhar, L. Malitsa, E. Mohammadi, D. Ronzani, and S. Capkun. Anonymous Communication for Messengers via "Forced" Participation. Technical report, available under `https://eprint.iacr.org/2017/191`, 2017.

[15] J. van den Hooff, D. Lazar, M. Zaharia, and N. Zeldovich. Vuvuzela: Scalable private messaging resistant to traffic analysis. In *Proceedings of the 25th ACM Symposium on Operating Systems Principles (SOSP 2015)*, Monterey, California, October 2015.

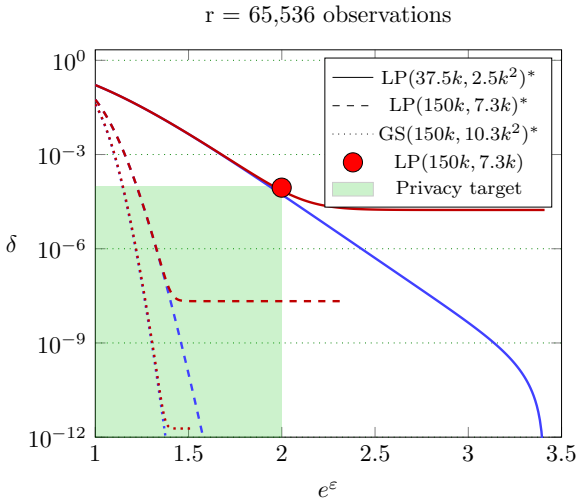


(a) After $r = 1,024$ observations with Laplace noise with $\mu = 4k$ and $\sigma = 330$ (solid), Laplace noise $\mu = 8k, \gamma = 500$ (dashed), and Gaussian noise with $\mu = 8k$ and $\sigma = 707$ (dotted), and the red dot represents the ε, δ combination for $\mu = 8k, \gamma = 500$ from the original Vuvuzela paper.

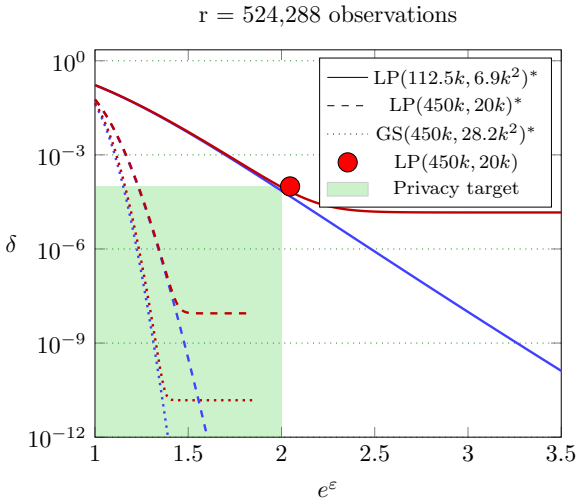


(b) After $r = 8,192$ observations with Laplace noise with $\mu = 10k$ and $\sigma = 827$ (solid), Laplace noise $\mu = 20k, \gamma = 1130$ (dashed), and Gaussian noise with $\mu = 20k$ and $\sigma = 1598$ (dotted), and the red dot represents the ε, δ combination for $\mu = 20k, \gamma = 1130$ from the original Vuvuzela paper.

Figure 13: The (ε, δ) graphs (y-axis and x axis, respectively, y-axis in $\log_{10}$-scale) from our method in comparison with the bound from the original Vuvuzela paper (for the dialing protocol). The figure depicts upper (red) and a lower bounds (blue) and Vuvuzela's privacy target (green, $\delta \leq 10^{-4}, e^{\varepsilon} \leq 2$).



(a) After $r = 65,536$ observations with Laplace noise with $\mu = 37.5k$ and $\sigma = 2.3k$ (solid), Laplace noise $\mu = 150k, \gamma = 7.3k$ (dashed), and Gaussian noise with $\mu = 150k$ and $\sigma = 10.3k$ (dotted), and the red dot represents the ε, δ combination for $\mu = 150k, \gamma = 7.3k$ from the original Vuvuzela paper.



(b) After $r = 524,288$ observations with Laplace noise with $\mu = 112.5k$ and $\sigma = 6.9k$ (solid), Laplace noise $\mu = 450k, \gamma = 20k$ (dashed), and Gaussian noise with $\mu = 450k$ and $\sigma = 28.2k$ (dotted), and the red dot represents the ε, δ combination for $\mu = 450k, \gamma = 20k$ from the original Vuvuzela paper.

Figure 14: The (ε, δ) graphs (y-axis and x axis, respectively, y-axis in $\log_{10}$-scale) from our method in comparison with the bound from the original Vuvuzela paper (for the dialing protocol). The figure depicts upper (red) and a lower bounds (blue) and Vuvuzela's privacy target (green, $\delta \leq 10^{-4}, e^{\varepsilon} \leq 2$).