Ratio Buckets: A Numeric Method for r-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

November 1, 2017

Abstract

Privacy guarantees of a privacy-enhancing system have to be robust against thousands of observations for many realistic application scenarios, such as anonymous communication systems, privacy-enhancing database queries, or privacy-enhancing machine-learning methods. The notion of r-fold Approximate Differential Privacy (ADP) offers a framework with clear privacy bounds and with composition theorems that capture how the ADP bounds evolve after r observations of an attacker. Previous work, however, provides privacy bounds that are loose, which results in an unnecessarily high degree of recommended noise, leading to low accuracy.

This work improves on previous work by providing upper and lower bounds for r-fold ADP, which enables us to quantify how tight our bounds are. We present a novel representation of pairs of distributions, which we call ratio buckets. We also devise a numerical method and an implementation for computing provable upper and lower bounds with these ratio buckets for ADP for a given number of observations. In contrast to previous work, our bucket method uses the shape of the probability distributions, which enables us to compute tighter bounds. Our studies indicate that previous work by Kairouz et al. provides tight bounds for the Laplace mechanism. However, we show that our work provides significantly tighter bounds for other scenarios, such as the Gaussian mechanism or for real-world timing leakage data. We show that it is beneficial 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, 10 times less noise already suffices, which significantly reduces Vuvuzela's overall bandwidth requirement.

[*] The authors are in alphabetical order. Both authors equally contributed to this work.

Contents

1 Introduction

Privacy analyses of privacy-enhancing systems, such as anonymous communication systems [18], privacy-enhancing database queries [4], and privacy-enhancing machine-learning methods [1], play a crucial role in understanding the effectiveness of these systems. The notion of *differential privacy* [4] and its important relaxation *approximate differential privacy* (written as (ε, δ)-ADP, or ADP for short [5, 15]) quantify, in terms of two parameters ε and δ, the privacy leakage against a strong worst-case attacker and have become an important tools for estimating privacy. In many application scenarios, the privacy has to hold under continual observation, i.e., against attackers that make thousands if not hundreds of thousands of observations. These ε and δ inevitably grow under continual observation, thus privacy eventually deteriorates (see Apple's case [17]).

Continual observation is formalized in a notion of r-fold ADP. Beside using randomness that is more robust under continual observation, the estimation of tight bounds for r-fold ADP can significantly contribute to the challenge of developing effective privacy-enhancing systems, i.e., the study of the question: "how much do ε and δ grow after r repeated observations?"

Since the initial r-fold ADP composition theorem [15], significant improvements were made in finding increasingly tighter bounds [7, 12]. While experiments indicate that some of these bounds [12] are tight w.r.t. the Laplace mechanism, this paper shows that for other mechanisms these generic results are not tight and substantially tighter bounds can be found, e.g., for the Gaussian mechanism or for estimating the privacy leakage of timing side-channels. Previous work that does not concentrate on one kind of noise lacks one major ingredient: it is oblivious to the underlying probability distributions and cannot, by its very nature, factor in the distribution-specific behavior under continual observation. Hence, results for r-fold ADP that are based on initial ε_0 and δ_0 values inherently assume a worst-case behavior under composition for those distributions.

1.1 Contribution

This work presents a numeric method for computing provable upper and lower bounds for r-fold ADP.

Our contribution is fourfold. First, we present a novel representation of the privacy leakage in the sense of ADP (i.e., worst-case privacy loss). The core idea of this work is to compute for a concrete pair of distributions (V, W) and all their atomic events x the quotient $V(x)/W(x)$ and to throw all atomic events with similar quotients into the same bucket.[1] We argue in Section 2 that such a pair of concrete distributions can be found in many cases by considering worst-case distributions. With these buckets we derive upper and lower bounds for r-fold ADP. We implement an iterative algorithm that computes these upper and lower bounds. The runtime complexity is dominated by the composition operation: $O(B^2)$, for a number of buckets B (determining the granularity of the approximation). We can optimize this computation in many cases, where the input distributions do not change from one observation to the next: via repeated squaring r-fold ADP only needs $\log_2(r)$ composition operations.

Second, we illustrate that our method is tighter than the previously best known generic bound: Kairouz et al.'s composition theorem [12]. As we were not able to directly compare the two results, we carefully implement an approximation of their bounds. While their bounds seem to be tight for the Laplace mechanism, we show that they are not tight for the Gaussian mechanism and for a model of timing-leakage measurements from the anonymous communication system CoverUp [16]. Using our lower bounds we show that our method is tight.

Third, we use our method to compare the Gaussian with the Laplace mechanism. We find that the r-fold ADP bounds for Laplace noise converges to the r-fold ADP for Gauss noise with half the variance of the Laplace noise.[2] We show that for the same variance, Gaussian noise provides significantly stronger privacy guarantees under a high number of observations.

Fourth, we illustrate the relevance of tighter bounds by improving the privacy analysis of the anonymity network Vuvuzela [18], which uses random noise to increase privacy. Vuvuzela declared a privacy goal of $\varepsilon = \ln 2$ and $\delta = 10^{-4}$. With the original Laplace noise, the tighter analysis shows that 2 to 4-fold reduced noise achieves the desired privacy goals, while with Gaussian noise already 5 to 10-fold reduced noise achieves the privacy goals.[3] If we do not reduce the amount of noise but keep the amount recommended in the Vuvuzela paper, the tighter analysis leads for Laplace noise a 3 to 4 orders of magnitude lower delta and for Gaussian noise with the same variance 4 to 6 orders of magnitude lower delta.

1.2 Discussion of our result

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

[1] This method is asymmetric, i.e., we have to compute it for $V(x)/W(x)$ and for $W(x)/V(x)$.

[2] We stress that this is not an example of the central limit theorem, which states that the sum of many independent random variables is normally distributed. The composition is not the convolution but the product of distributions.

[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.

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 extent 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 little difference in the upper and lower bound.

While our result expects concrete distributions as input, we show that in many cases concrete worst-case distributions can be found for k-fold ADP 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 r-fold ADP 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.

A disclaimer: this work constitutes a powerful tool for finding tight bounds in a r-fold ADP privacy analysis of a privacy-enhancing system; devising the privacy analysis is outside of the scope of this work, e.g., finding the right level of abstraction, a useful attacker model, suitable usage behaviors, the right privacy mechanism.

1.3 Worst case distributions for Differential Privacy

Classically, differential privacy is defined for all pairs of neighboring databases. 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. In practice, however, there is a direct connection between the worst-case choices of scenarios or adversarial decisions and very simple concrete distributions. Let us consider counting queries q with sensitivity 1 to which Laplace noise is added: the mechanism M that gets a database D as input is defined as $M(D) := q(D) + \mathrm{LP}_0$, where LP_i is the Laplace distribution with mean i. In this example, it suffices to only consider LP_0 and LP_1 with means 0 and 1, instead of considering $M(D_0)$ and $M(D_1)$ for all possible combinations of neighboring databases D_0 and D_1. Let D_0 and D_1 be two such neighboring databases where the true answers to a query q are $q(D_0) = x$ and $q(D_1) = x + 1$, respectively, for some x. We can map any output y drawn from LP_i (for $i \in \{0, 1\}$) to $y + x$ to obtain the correct adversarial view for the respective scenario $M(D_i) = q(D_i) + \mathrm{LP}_0$. In general, to formally apply our approach we need two distributions and the existence of a reduction: given a description of the scenario or an adversarial choice as well as an output of one of the distributions we consider in our calculation, the reduction produces the respective output within the differential privacy scenario. In other words, there is an efficient permutation that translates the distributions we analyze into the distributions of the respective scenario.

2 Related work

Differential privacy Differential Privacy (DP) [4] quantifies how closely related two similar distributions are from an information-theoretic perspective. In case 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, 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 $\forall S \subseteq \mathcal{U}.\ P_A(x \in S|x) \leq e^{\varepsilon} \cdot P_B(x \in S|x)$ (and vice versa). To extend the applicability of DP, approximate differential privacy [5] (ADP) allows for distributions to exceed a limiting factor ε, as long as this is in total only exceeded by a small value δ. Formally, we say that two distributions A and B over the universe $\mathcal{U}$ are (ε, δ)-ADP, if $\forall S \subseteq \mathcal{U}.\ P_A(x \in S|x) \leq e^{\varepsilon} \cdot P_B(x \in S|x) + \delta$ (and vice versa). The notion of *computational differential privacy* [15] replaces the sets S of events by adversarial distinguisher machines.

Generic bounds for DP under continual observation There ha been significant improvements on the original composition result. Using concentration bounds Dwork et al. [7] proved the advanced composition result for adaptive r-fold composition, which was further improved by Kairouz et al. [12]. These generic bounds have in common that they are oblivious to the actual distributions and their bounds only rely on the

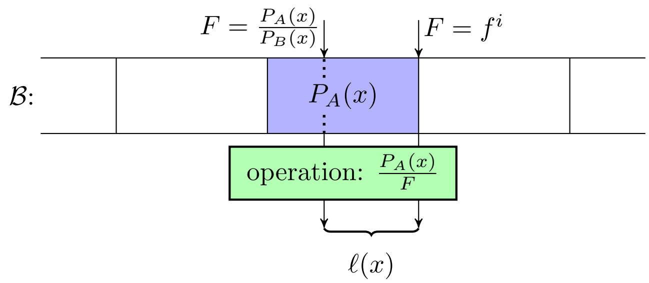


Figure 1: Depiction of how an element x is placed into a bucket. Buckets store f^i and $P_A(x)$ (accumulated over all elements in the bucket). We approximate $P_B(x)$ with $\frac{P_A(x)}{f^i}$, accepting an error of $\ell(x) = P_B(x) - \frac{P_A(x)}{f^i}$.

initial values $(\varepsilon_0, \delta_0)$; hence, their optimality only holds for the worst-case distribution with the given $(\varepsilon_0, \delta_0)$ (one-shot DP). We show in this paper, that retaining the shape of the distributions can lead to significantly tighter bounds.

Adaptive mechanisms For some applications there has been work on adaptive DP-mechanisms that achieve differential privacy under continued observation by using carefully correlated noise and only using noise when necessary [1, 6, 8, 10, 11]. While distributions over correlated noise cannot directly be input to our bucket approach, the proofs of these adaptive mechanisms can still benefit from our results as they often over-approximate a subset of these correlated distributions with independent distributions, e.g., in order to apply Azuma's Inequality [10] (which is stated for independent distributions). A concrete example where an adaptive mechanism can benefit from our approach is the of Abadi et al. [1] that uses Gaussian noise for privacy preserving deep learning. While their approach takes the shape of the distribution into account to some degree and their bounds are significantly tighter than previous work, our bucket approach can achieve even tighter bounds. We refer to Appendix A.1 for an example calculation.

Dependencies The work of Liu, Chakraborty and Mittal [14] 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 in DP's basic definition. Their approach, however, finally results in the same techniques as in previous work 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 Recent work [9, 13] made progress on finding optimal mechanisms for DP for a large class of utility functions. These results concentrate on single observations and do not characterize how these mechanism behave under k-fold composition.

3 Ratio buckets of two distributions

3.1 Informal description of ratio buckets

Generic bounds for differential privacy under continual observation [7, 12] are stated independently of the shape of the underlying distributions, simply based on the ADP guarantees before the composition. This obliviousness is both the greatest strength and the greatest weakness of these generic bounds. The exact shape of the distribution does not need to be characterized to apply these results, but they cannot devise tight bounds that are derivable from the shape of the distributions.

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 2: 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 $f^{i-1} P_B(x) \leq P_A(x) \leq f^i P_B(x)$, the bucket $\mathcal{B}_A(-n)$ contains all elements with $P_A(x) \leq f^{-n} P_B(x)$ and the bucket $\mathcal{B}_A(\infty)$ contains all elements with $P_A(x) > f^n P_B(x)$.

$P_A(x)$	the prob. that x happens in A.
ε, δ	parameters for DP.
$\mathcal{U}$	universe of all events.
f	factor (close to 1) with $f > 1$.
n	determines number of buckets.
∞	symbol for any ratio $> f^n$.
$\mathcal{B}(A, B, f, n)$	A/B ratio buckets with indexes $\{-n, \ldots, n, \infty\}$ and ratios $\{\leq f^{-n}, \ldots, \leq f^n, > f^n\}$.
$\mathcal{B}(i)$ for $i \in N$	bucket with index i.
$\mathcal{B}(x)$ for $x \in \mathcal{U}$	impact of the event x.
$\ell, \ell(i), \ell(x)$	our "real" error correction term.
$\tilde{\ell}, \tilde{\ell}(i), \tilde{\ell}(x)$	bound on the maximum error, "virtual" error correction term.
$\iota_B(x)$	index of x in ratio buckets B.

Figure 3: Notation for our ratio buckets.

We now introduce an alternative approach: we approximate the distributions and the ways they are related. Given two distributions A and B, the most important feature (for differential privacy) of each event x is the ratio between the probability of x in A, denoted by $P_A(x)$, and the probability of x in B, denoted by $P_B(x)$. We group events by this ratio $\frac{P_A(x)}{P_B(x)}$. Then, to render this approach feasible, we collect all such x with a similar factor into the same set, which we call a *bucket*. Given a factor $f > 1$, the bucket $\mathcal{B}(i)$ summarizes all events where $f^{i-1} < \frac{P_A(x)}{P_B(x)} \leq f^i$ (illustrated in Figure 1). The value of $\mathcal{B}(i)$ is the sum over the probabilities $P_A(x)$ of all those events (according to distribution A). We approximate $\sum_x P_B(x)$ in $\mathcal{B}(i)$ as $\mathcal{B}(i)/f^i$, thereby introducing for each $P_B(x)$ an error of $\ell(x) = P_B(x) - \frac{P_A(x)}{f^i}$.

This buckets representation is suited for composing two pairs of distributions, say (A_1, B_1) with (A_2, B_2). We combine buckets $\mathcal{B}_1(i)$ and $\mathcal{B}_2(j)$ multiplicatively and obtain the probability of all events in a new bucket $(\mathcal{B}_1 \times \mathcal{B}_2)(i+j)$, for which we still have $\frac{P_{A_1}(x)}{P_{B_1}(x)} \cdot \frac{P_{A_2}(x)}{P_{B_2}(x)} \leq f^{i+j}$ (c.f. Figure 5).

3.2 Differential privacy

We review the definition for approximate differential privacy (ADP), generalized to a pair of distributions. ADP characterizes privacy by a multiplicative value ε and an additive error value δ. In particular, we introduce *tight ADP* to characterize for a given ε the smallest values of δ for which ADP is satisfied.

Definition 1 ((Tight) ADP)**.** *Two distributions A and B over the universe $\mathcal{U}$ are (ε, δ)-ADP, if $\forall$ sets*

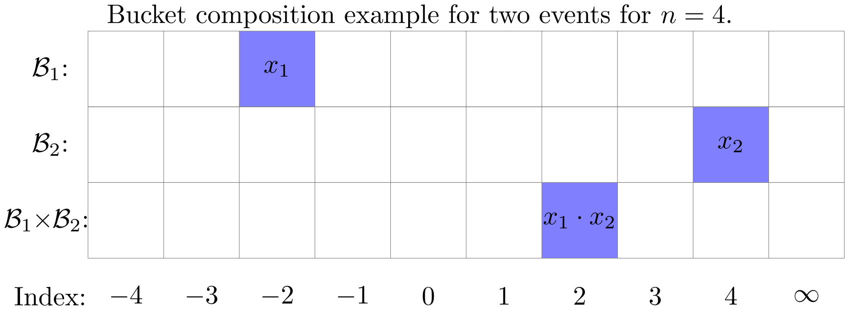

Bucket composition example for two events for $n = 4$.
$\mathcal{B}_1$:
x_1
$\mathcal{B}_2$:
x_2
$\mathcal{B}_1 \times \mathcal{B}_2$:
$x_1 \cdot x_2$
Index: -4 -3 -2 -1 0 1 2 3 4 ∞

Figure 4: Depiction of how individual events x_1 with index -2 and x_2 with index 4 compose into their new bucket with index $-2 + 4 = 2$.

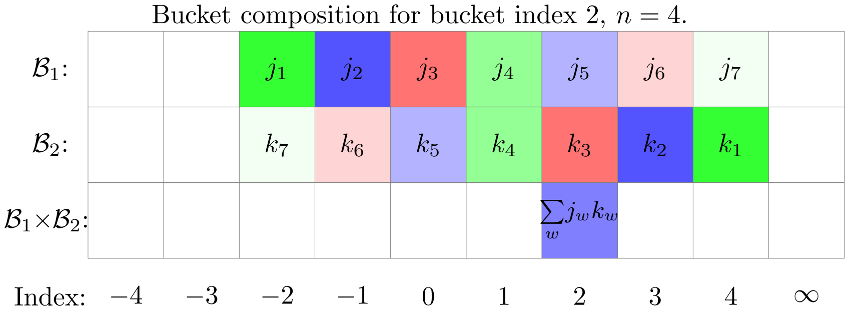

Bucket composition for bucket index 2, $n = 4$.
$\mathcal{B}_1$:
j_1 j_2 j_3 j_4 j_5 j_6 j_7
$\mathcal{B}_2$:
k_7 k_6 k_5 k_4 k_3 k_2 k_1
$\mathcal{B}_1 \times \mathcal{B}_2$:
$\sum_w j_w k_w$
Index: -4 -3 -2 -1 0 1 2 3 4 ∞

Figure 5: Depiction of the bucket composition for the (new) bucket with index $i = 2$. We calculate the value of bucket i by summing over the product of all $\mathcal{B}_1(j_w) \cdot \mathcal{B}_2(k_w)$. Graphically, buckets with the same color are combined. Note that none of the buckets $\infty, -3$ and -4 are used for the composition, as for all $j \in \{-4, \dots, 4\}$, $\infty + j \neq 2$, $-3 + j \neq 2$ and $-4 + j \neq 2$.

$S \subseteq \mathcal{U}$,

$$P_A(S) \leq e^{\varepsilon} P_B(S) + \delta \text{ and}$$
$$P_B(S) \leq e^{\varepsilon} P_A(S) + \delta.$$

A and B are tightly *(ε, δ)-ADP if they are (ε, δ)-ADP, and $\forall \delta' < \delta$, A and B are not (ε, δ')-ADP.* [4]

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

Lemma 1. *For every ε, two distributions A and B over a finite universe $\mathcal{U}$ are tightly (ε, δ)-ADP with*

$$\delta = \max\left(\sum_{x \in U} \max\left(P_A(x) - e^{\varepsilon} P_B(x), 0\right),\right.$$
$$\left.\sum_{x \in U} \max\left(P_B(x) - e^{\varepsilon} P_A(x), 0\right)\right)$$

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 $\forall S \subseteq \mathcal{U}$,

$$P_A(x \in S : x) - e^{\varepsilon} P_B(x \in S : x)$$
$$\leq \sum_{x \in U} \max\left(P_A(x) - e^{\varepsilon} P_B(x), 0\right)$$

[4] Our bucket approach can also be used to find a bound for ε-tightness for a given δ, i.e., $\forall \varepsilon' < \varepsilon$, A and B are not (ε', δ)-ADP.

The inverse direction then follows analogously.

$$\begin{aligned}
&P_A(x \in S : x) - e^{\varepsilon} P_B(x \in S : x)\\
=&\sum_{x \in S} P_A(x) - e^{\varepsilon} P_B(x)\\
\leq&\sum_{x \in S} \max\left(P_A(x) - e^{\varepsilon} P_B(x), 0\right)\\
\leq&\sum_{x \in U} \max\left(P_A(x) - e^{\varepsilon} P_B(x), 0\right)
\end{aligned}$$

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

$$\begin{aligned}
&P_A(x \in S : x) - e^{\varepsilon} P_B(x \in S : x)\\
=&\sum_{x \in S} P_A(x) - e^{\varepsilon} P_B(x)\\
=&\sum_{x \in U} \max\left(P_A(x) - e^{\varepsilon} P_B(x), 0\right).
\end{aligned}$$

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

$$\begin{aligned}
&P_B(x \in S : x) - e^{\varepsilon} P_A(x \in S : x)\\
=&\sum_{x \in S} P_B(x) - e^{\varepsilon} P_A(x)\\
=&\sum_{x \in U} \max\left(P_B(x) - e^{\varepsilon} P_A(x), 0\right).
\end{aligned}$$

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 It might appear preferable to only include the distinguishing events into δ, to guarantee pure ε-DP with probability $(1 - \delta)$. However, ε then inherently grows linearly in the number of observations r. Allowing, say, $100,000$ observations, results in a factor of $e^{100,000\varepsilon}$, which might deteriorate the results more than accepting a small increase in δ. An example of such a trade-off is depicted in Figure 15b.

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 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 $\forall 1 \leq i \leq k.\ ||x_i - y_i||_1 \leq s$:*

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

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 in 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 (in the sense of Section 1.3) 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, 12]. 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 $\hat{\delta} = k \cdot \delta$ 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 [12])**.** *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, these results 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.

3.4 Ratio buckets

We approximate the features underlying the pair of distributions in a way that is sufficient for calculating (ε, δ)-ADP, the *ratio buckets*, and that comes with an efficient way for computing the k-fold (ε, δ)-ADP from a sequence of ratio buckets.

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 lists. 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 list Given two distributions A and B, the idea is to group all atomic events of A and B by the ratio between their probabilities in A and in B: if x is an atomic event from the universe $\mathcal{U}$, we consider $\frac{P_A(x)}{P_B(x)}$ and, depending on this value, decide in which set we put the event. If this fraction is undefined because $P_B(x) = 0$, we put the event in a specific set.

Given two system parameters f and n, we then summarize these groups into $2n + 2$ sets (the *buckets*), where f is the separation-factor f and n the limit. For $i \in \{-n, \dots, n\}$ we assign an elementary event x to the set S_i, if $P_A(x) \leq f^i P_B(x)$ and if for all $j \in \{-n, \dots, n\}$ with $j < i$ we have $P_A(x) > f^j P_B(x)$ (c.f. Figure 1). All remaining events with $f^n P_B(x) < P_A(x)$ are assigned to the special set S_∞. For each set S_i (including S_∞), we accumulate the $P_A(x)$ of all events $x \in S_i$ to obtain the respective *bucket* $\mathcal{B}(i)$. After creating the buckets, we do not keep any information about the atomic events or the sets S_i; all further calculations are based on the $2n+2$ buckets $\mathcal{B}(-n), \dots, \mathcal{B}(n), \mathcal{B}(\infty)$, the *bucket list.* Thus, the runtime of all further calculations only depends on the number of buckets. After composing two such ratio buckets with each other, we yield another ratio bucket with the same parameters for f and n. Although the precision of our method may decrease with the number of compositions, the complexity of all operations remains the same.

Our bucket approach is asymmetric as we consider the probabilities of events occurring in A in relation to the probabilities of the same events occurring in B and over-approximate the factor between them slightly (c.f. Figure 1). Thus, A/B ratio buckets only deliver guarantees on one direction of differential privacy – in practice we create both direction, A/B and B/A ratio buckets.

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/B* ratio buckets $\mathcal{B}$ *over the universe $\{-n, -n+1, \dots, n\} \cup \{\infty\}$ s.t.*

$$\forall i \in \{-n, \dots, n\} \cup \{\infty\} . \mathcal{B}(i) = \textstyle\sum_{x \in S_i} P_A(x),$$

where the sets S_i are defined as follows:

$$\begin{aligned} S_\infty &= \{x \in \mathcal{U}. P_A(x) > f^n P_B(x)\} \\ \forall i \in \{-n+1, \dots, n\}\, S_i &= \left\{x \in \mathcal{U}.\ f^{i-1} P_B(x) < P_A(x) \leq f^i P_B(x)\right\} \\ S_{-n} &= \left\{x \in \mathcal{U}. P_A(x) \leq f^{-n} P_B(x)\right\}. \end{aligned}$$

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

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

We next define ADP directly on a bucket list. For all events x in $S_i \neq S_\infty$, we know that $P_A(x) \leq f^i P_B(x)$. We perform a slight over-approximation by treating this inequality as an equality and then use $P_A(x) - P_A(x)/f^i$ as in Lemma 1. For $x \in S_\infty$, we add $P_A(x)$ to δ, counting them as total privacy-breakdowns.

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

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

Computing the composition of ratio buckets We proceed by defining how to compose ratio buckets. The composition operation shows the 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 4). Since events x_1 in $\mathcal{B}_1(i)$ satisfy $P_{A_1}(x_1) \leq f^i P_{B_1}(x)$ and events x_2 in $\mathcal{B}_2(j)$ satisfy $P_{A_2}(x_2) \leq f^j P_{B_2}(x)$, we know that the combined events (x_1, x_2) satisfy

$$\begin{aligned} &P_{A_1 \times A_2}((x_1, x_2)) \\ =&P_{A_1}(x_1) \cdot P_{A_2}(x_2) \\ \leq& f^i P_{B_1}(x_1) \cdot f^j P_{B_2}(x_2) \\ =& f^{i+j} P_{B_1 \times B_2}((x_1, x_2)). \end{aligned}$$

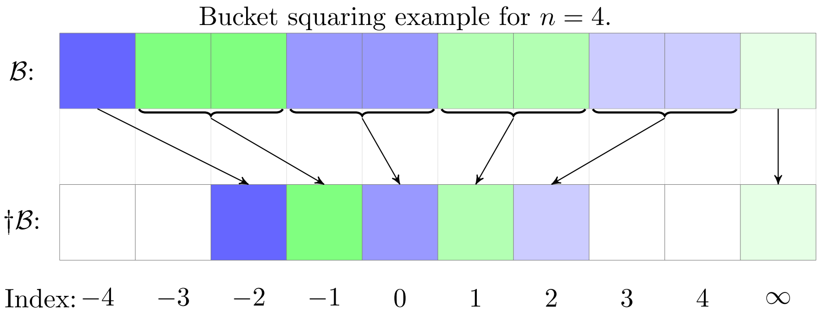

Bucket squaring example for $n = 4$.
$\mathcal{B}$:
$\dagger\mathcal{B}$:
Index: -4 -3 -2 -1 0 1 2 3 4 ∞

Figure 6: 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.

Following this strategy, we can hence maintain the desired property $P_A(x) \leq f^i P_B(x)$ for all events in bucket i, even after composition We refer to Figure 5 for a graphical depiction of the bucket composition for one bucket. More formally, we define the composition of two ratio buckets 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 A_1/B_2 and A_2/B_2 ratio buckets (resp.). We define the composition of the pairs as $\mathcal{B}_1 \times \mathcal{B}_2$, where* $\forall i \in \{-n, -n+1, \ldots, n\} \cup \{\infty\}$,

$$\mathcal{B}_1 \times \mathcal{B}_2(i) := \begin{cases} \sum_{j+k=i} \mathcal{B}_1(j) \cdot \mathcal{B}_2(k) & i \in N \setminus \{-n\} \\ \sum_{j+k \leq -n} \mathcal{B}_1(j) \cdot \mathcal{B}_2(k) & i = -n \\ \sum_{j+k > n} \mathcal{B}_1(j) \cdot \mathcal{B}_2(k) & i = \infty \end{cases}$$

We stress that as the sets S_i for $i \in \{-n, \ldots, n\} \cup \{\infty\}$ describe a partitioning of $\mathcal{U}$ and the buckets $\mathcal{B}_1$ and $\mathcal{B}_2$ add up to 1, i.e,

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

When composing ratio buckets, the bucket list naturally "broadens", i.e., the buckets that are farther away from the middle bucket (with factor f^0) gain higher values. When creating ratio buckets 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 the value of f, the more precise the ratio buckets model of the features of the distributions, but the fewer compositions before a significant amount of events reaches the corner buckets $\big(\mathcal{B}(-n)$ and $\mathcal{B}(\infty)\big)$, which again reduces the precision. To counter this effect, we introduce an additional operation which we call *squaring*: we square the factor f, thus halving the precision of the ratio buckets, and merge the ratio buckets into these new, more coarse-grained ratio buckets. Squaring allows us to start with much more fine-grained ratio buckets and reduce the granularity as we compose, which can significantly improve the overall precision of the approach. We choose to square f instead of increasing it to an arbitrary f' to ease the computation of the new ratio buckets: 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 6 for a graphical depiction of squaring.

$$\dagger\mathcal{B}_1(i) := \begin{cases} \mathcal{B}_1(2i-1) + \mathcal{B}_1(2i) & i \in [-n/2+1, n/2] \\ \mathcal{B}_1(\infty) & i = \infty \\ 0 & \textit{otherwise} \end{cases}$$

The composition of ratio buckets is commutative but not associative. Moreover, when and how often the squaring was performed influence the resulting ratio buckets. Hence, we need to keep track of the order in which we applied composition and squaring. To this end, we define *composition trees.* These 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* *labeled with a bucket factor $f > 1$; leaves ($T = \varnothing(A_i, B_i)$) are additionally labeled 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 $\varnothing(A_i, B_i)$. 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 ratio buckets 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 ratio buckets that have 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 let $n \in \mathbb{N}$. 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 $k \in \{1, \ldots, \mathcal{W}\}$, $\iota_{\varnothing(A_k,B_k)}(x_k) := i$ where*

$$i = \begin{cases} l & \text{if } l \in \{-n+1, \ldots, n\} \wedge \\ & f^{l-1} P_{B_k}(x_k) < P_{A_k}(x_k) \le f^l P_{B_k}(x_k) \\ \infty & \text{if } P_{A_k}(x_k) > f^n P_{B_k}(x_k) \\ -n & \text{otherwise} \end{cases}$$

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

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

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

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

We stress that $\iota_{T_1 \times T_2}(x_1, x_2)$ 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

$$\iota_{(T_1 \times T_2) \times T_3}(x_1, x_2, x_3) \ne \iota_{T_1 \times (T_2 \times T_3)}(x_1, x_2, x_3).$$

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

We begin by showing a helpful lemma that directly follows our main strategy: all atomic events x that are assigned an index $i \ne \infty$ (according to a composition tree T) satisfy $P_A(x) \le f^i P_B(x)$.

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 $n \in \mathbb{N}$, 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 $\iota_T(x) \ne \infty$ and the root node has f in the label, we have $P_A(x) \le f^{\iota_T(x)} P_B(x)$. Analogously, with $\iota_{T^B}(x) \ne \infty$ we have $P_B(x) \le f^{\iota_T(x)} P_A(x)$.*

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

Let $x = (x_1, \dots, x_{\mathcal{W}}) \in \prod_{k=1}^{\mathcal{W}} \mathcal{U}_k$. Note that if $\iota_T(x) \neq \infty$, then for all $k \in \{1, \dots, \mathcal{W}\}$ $\iota_{\oslash(A_k,B_k)}(x_k) \neq \infty$.

For each k such that $\iota_{\oslash(A_k,B_k)}(x_k) = -n$, from the case distinction of $\iota_{\oslash(A_k,B_k)}(x_k)$ in Definition 7 it follows that

$$P_{A_k}(x_k) \leq f^{-n} P_{B_k}(x_k) \tag{1}$$

$$P_{A_k}(x_k) \leq f^{\iota_{\oslash(A_k,B_k)}(x_k)} P_{B_k}(x_k). \tag{2}$$

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

$$P_{A_k}(x_k) \leq f^{\iota_{\oslash(A_k,B_k)}(x_k)} P_{B_k}(x_k). \tag{3}$$

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

$$\begin{aligned} P_A(x) &= \prod_{k=1}^{\mathcal{W}} \underbrace{P_{A_k}(x_k)}_{\leq f^{\iota_{\oslash(A_k,B_k)}(x_k)} P_{B_k}(x_k)} \\ &\leq \prod_{k=1}^{\mathcal{W}} f^{\iota_{\oslash(A_k,B_k)}(x_k)} P_{B_k}(x_k) \\ &= \underbrace{\prod_{k=1}^{\mathcal{W}} f^{\iota_{\oslash(A_k,B_k)}(x_k)}}_{\leq f^{\iota_T(x)}} \underbrace{\prod_{k=1}^{\mathcal{W}} P_{B_k}(x_k)}_{= P_B(x)} \\ &\leq f^{\iota_T(x)} P_B(x) \end{aligned}$$

Note that $\sum\limits_{k \in \{1,\dots,\mathcal{W}\}} \iota_{\oslash(A_k,B_k)} \leq \iota_T(x)$ 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 labeled with f (and the squaring node thus is labeled with f^2), we know that $\iota_T(x) \neq \infty \equiv \iota_{T_1}(x) \neq \infty$. For $\iota_{T_1}(x) \neq \infty$ we know by the induction hypothesis that $P_A(x) \leq f^{\iota_{T_1}(x)} P_B(x)$. By definition, we have

$$\begin{aligned} P_A(x) \leq f^{\iota_{T_1}(x)} P_B(x) \qquad &= f^{2\iota_{T_1}(x)/2} P_B(x) \\ \leq (f^2)^{\lceil \iota_{T_1}(x)/2 \rceil} P_B(x) &= (f^2)^{\iota_T(x)} P_B(x) \end{aligned}$$

□

Lemma 7 (Bucket 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 $k \in \{1, \dots, \mathcal{W}\}$ $\mathcal{B}(A_k, B_k, f_k, n) = \mathcal{B}_k$ be A_k/B_k ratio buckets and let T be a composition tree. Let $\varepsilon \geq 0$. Let $\mathcal{B}_A := \prod_{k \in \{1,\dots,\mathcal{W}\}}^{T} \mathcal{B}(A_k, B_k, f_k, n)$. Then, $\mathcal{B}(i) = \sum_{x \text{ s.t. } \iota_T(x)=i} P_A(x)$.*

Proof. We show the lemma via structural induction over T. We only show the lemma for A, but the proof follows analogously for B. Let $N = \{-n, \dots, n\}$.

If $T = \oslash(A_i, B_i)$: Let $i \in N \cup \{\infty\}$. By Definitions 3 and 7 with S_i as in Definition 3,

$$\mathcal{B}(i) = \textstyle\sum_{x \in S_i} P_A(x) = \sum_{x, \iota(x)=i} P_A(x).$$

Otherwise, assume the lemma holds for ratio buckets $\mathcal{B}_1$ over a universe $\mathcal{U}_1$ and $\mathcal{B}_2$ over a universe $\mathcal{U}_2$ with composition trees T_1 and T_2 over distributions A_1, B_2 and A_2, B_2. Let $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2$ and $A = A_1 \times A_2$. We have for $i \in N \setminus \{-n\}$

```
BucketDelta($A, B, f, n, t, \varepsilon$):
$\mathcal{B}(-n) = \ldots = \mathcal{B}(n) = \mathcal{B}(\infty) = 0$
for $x \in \mathcal{U}$ do
    if $\exists i < n.\ P_A(x) \leq f^i P_B(x)$ then
        $\mathcal{B}(\max(-n, i))$ += $P_A(x)$
    else
        $\mathcal{B}(\infty)$ += $P_A(x)$
for $i$ from $0$ to $t$ do
    $\mathcal{B}' = \mathcal{B} \times \mathcal{B}$
    if $\mathcal{B}'(\infty) > 2.2 \cdot \mathcal{B}(\infty)$ then
        $\mathcal{B} = \dagger\mathcal{B}$
    $\mathcal{B} = \mathcal{B} \times \mathcal{B}$
return $\delta(\mathcal{B}, \varepsilon)$
```

Figure 7: Depiction of how we create buckets – for simplicity without error correction terms and for the common special case where we compose the same distributions ($A_1 = A_2 = \ldots = A_r$ and $B_1 = B_2 = \ldots = B_r$). We use repeated squaring to compute r-fold DP for $r = 2^t$ compositions.

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

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

$$\begin{aligned}
&= \sum_{x=(x_1,x_2) \in \mathcal{U}_1 \times \mathcal{U}_2 \ s.t. \ \iota_T(x)=i} \mathcal{B}_1(x_1) \cdot \mathcal{B}_2(x_2) \\
&\overset{\text{Definition 7}}{=} \sum_{x=(x_1,x_2) \in \mathcal{U} \ s.t. \ \iota_T(x)=i} \mathcal{B}(x).
\end{aligned}$$

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 $\iota_T(x) = -n$ is equivalent to $\iota_{T_1}(x_1) + \iota_{T_2}(x_2) \leq -n$ and for ∞ we have $j + k > n$ and we know from Definition 7 that $\iota_T(x) = \infty$ is equivalent to $\iota_{T_1}(x_1) + \iota_{T_2}(x_2) \geq n$.

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

$$\dagger\mathcal{B}_1(i) = 0 = \textstyle\sum_{x \in \emptyset} P_{A_1}(x) = \sum_{x \in \mathcal{U}_1, \iota_{\dagger T_1} = i} P_{A_1}(x).$$

For $i = \infty$, we have $\dagger\mathcal{B}_1(\infty) = \mathcal{B}_1(\infty)$, so the statement follows from the IH. For $i \in \{-n/2 + 1, \ldots, n/2\}$ we have

$$\begin{aligned}
&\dagger\mathcal{B}_1(i) = \mathcal{B}_1(2i) + \mathcal{B}_1(2i - 1) \\
&\overset{\text{IH}}{=} \textstyle\sum_{x \in \mathcal{U}_1, \iota_{T_1}(x)=2i} P_{A_1}(x) + \sum_{x \in \mathcal{U}_1, \iota_{T_1}(x)=2i-1} P_{A_1}(x) \\
&= \textstyle\sum_{x \in \mathcal{U}_2 \iota_T(x)=i} P_{A_1}(x).
\end{aligned}$$

The statement for $\dagger\mathcal{B}_1(-n/2)$ follows analogously. □

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

Theorem 1 (Buckets are sound)**.** *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 $k \in \{1, \ldots, \mathcal{W}\}$ $\mathcal{B}(A_k, B_k, f_k, n) = \mathcal{B}_k$ be ratio buckets and let T be a composition tree. Let $\varepsilon \geq 0$. Moreover, let $\mathcal{B}_A := \prod_{k\in\{1,\ldots,\mathcal{W}\}}^{T} \mathcal{B}(A_k, B_k, f_k, n)$ and analogously $\mathcal{B}_B := \prod_{k\in\{1,\ldots,\mathcal{W}\}}^{T} \mathcal{B}(B_k, A_k, f_k, n)$.*

If for $\varepsilon, \delta \geq 0$, $\mathcal{B}_A$ is (ε, δ)-ADP and $\mathcal{B}_B$ is (ε, δ)-ADP, then A and B are (ε, δ)-differentially private.

The theorem follows quite trivially from the proof of Lemma 13 in the subsequent chapter. We still present the proof as it could be helpful in understanding the soundness of our ratio buckets.

Proof. We show the proof for $\mathcal{B}_A$; the proof for $\mathcal{B}_B$ follows analogously.Let $N = \{-n, \ldots, n\}$. By definition, $\mathcal{B}_A$ is (ε, δ)-ADP if

$$\delta \geq \sum_{i\in N} \left(\max\left(0, \mathcal{B}(i) \cdot (1 - e^{\varepsilon}/f^{i})\right)\right) + \mathcal{B}(\infty).$$

We ignore $\mathcal{B}(\infty)$ for now and apply Lemma 7 and get

$$\begin{aligned} &\sum_{i\in N} \left(\max\left(0, \sum_{x\in\mathcal{U}.\iota_T(x)=i} P_A(x) \cdot (1 - \tfrac{e^{\varepsilon}}{f^{i}})\right)\right) \\ =&\sum_{i\in N.f^{i}>e^{\varepsilon}} \left(\sum_{x\in\mathcal{U}.\iota_T(x)=i} P_A(x) \cdot (1 - \tfrac{e^{\varepsilon}}{f^{i}})\right) \end{aligned}$$

Using Lemma 6 we get

$$\sum_{x\in\mathcal{U}.\iota_T(x)=\in N.f^{\iota_T(x)}>e^{\varepsilon}} \max\left(0, P_A(x) - e^{\varepsilon} P_B(x)\right).$$

With $\mathcal{B}(\infty)$ (where we also apply Lemma 7) we yield

$$\begin{aligned} &\sum_{x\in\mathcal{U}.\iota_T(x)\in N.f^{\iota_T(x)}>e^{\varepsilon}} \max\left(0, P_A(x) - e^{\varepsilon} P_B(x)\right) \\ &+\sum_{x\in\mathcal{U}.\iota_T(x)=\infty} P_A(x) \\ \geq&\sum_{x\in\mathcal{U}} \max\left(0, P_A(x) - e^{\varepsilon} P_B(x)\right). \end{aligned}$$

We repeat the calculation analogously for B and using Lemma 1 we see that A and B are indeed (ε, δ)-ADP. □

4 Reducing and bounding the error

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

There are two reasons for why events can be misplaced: First, when composing ratio buckets, events can be misplaced by one bucket. We take care of this by not including the EC 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 EC 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. Assume two distributions A and B: Whenever we enter an event x into a bucket $\mathcal{B}(i)$, we store 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: $\ell(i) \mathrel{+}= P_B(x) - \frac{P_A(x)}{f^i}$. 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* (or EC terms) per individual bucket. We refer to Figure 1 for a graphical intuition of our error correction.

As an example consider one bucket $\mathcal{B}(i)$, containing events $x \in S_i$ for a set S_i:

$$\begin{aligned}\frac{\mathcal{B}(i)}{f^i} - \ell(i) =& \frac{\sum_{x \in S_i} P_A(x)}{f^i} \\ &- \sum_{x \in S_i} \left(P_B(x) - \frac{P_A(x)}{f^i} \right) \\ =& \sum_{x \in S_i} P_B(x).\end{aligned}$$

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 omit the EC 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 that bucket, namely the ones with $P_A(x) \leq e^\varepsilon P_B(x)$, 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 EC 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 EC term.

Definition 8 (Ratio buckets with error correction terms)**.** *Let A, B be distributions over the universe $\mathcal{U}$, let $f > 1$ and $n \in \mathbb{N}$ and let $N = \{-n, \dots, n\}$. We define A/B ratio buckets with EC terms $\mathcal{B}(A, B, f, n) = (\mathcal{B}, \tilde{\ell}, \ell, f, 1)$, as $\forall i \in N \cup \{\infty\}$*

$$\begin{aligned}\mathcal{B}(i) :=& \textstyle\sum_{x \in \mathcal{U} \ s.t. \ \iota(x) = i} P_A(x) \\ \tilde{\ell}(i) :=& \begin{cases} \sum_{x \in \mathcal{U}, \iota(x) = i} P_B(x) - \frac{P_A(x)}{f^i} & i \in N \\ 0 & i = \infty \end{cases} \\ \ell(i) :=& \begin{cases} \tilde{\ell}(i) & i \in N \setminus \{-n\} \\ 0 & i \in \{-n, \infty\} \end{cases}\end{aligned}$$

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: For the composition, we want to calculate the error correction (EC) term for the combined events: given events x_1 and x_2 with (individual) error terms $P_{B_1}(x_1) - \frac{P_{A_1}(x)}{f^{\iota_1}}$ and $P_{B_2}(x_2) - \frac{P_{A_2}(x_2)}{f^{\iota_2}}$ we want (in the typical case, ignoring corner cases) to have an EC term for the pair of the form $P_{B_1 \times B_2}((x_1, x_2)) - \frac{P_{A_1 \times A_2}((x_1, x_2))}{f^{\iota_1 + \iota_2}}$. However, the buckets cannot keep track of the value for $P_{B_1 \times B_2}((x_1, x_2))$– recall that this is precisely why we have introduced the error terms. Fortunately, we can calculate the desired EC terms from the previous EC terms ℓ_1, ℓ_2, the bucket values $\mathcal{B}_1, \mathcal{B}_2$, and the bucket factor f as

$$\ell(i) := \textstyle\sum_{j+k=i} \frac{\mathcal{B}_1(j)}{f^j} \ell_2(k) + \frac{\mathcal{B}_2(k)}{f^k} \ell_1(j) + \ell_1(j) \ell_2(k).$$

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 their EC 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 EC terms need to be modified to capture this change in the bucket factor, based on the previous EC terms ℓ_1 and bucket values $\mathcal{B}_1$:

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

We define the composition and squaring as follows.

Definition 9 (Composition and squaring with EC terms)**.** *For ratio buckets* $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ *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* $f_1 = f_2 = f$*, let* $N = \{-n, \ldots, n\}$*. We have*

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

In all of the following sums we sum over all $j, k \in N$*.*

$$\mathcal{B}_1 \times \mathcal{B}_2(i) := \begin{cases} \sum_{j+k=i} \mathcal{B}_1(j) \cdot \mathcal{B}_2(k) & i \in N \setminus \{-n\} \\ \sum_{j+k \le -n} \mathcal{B}_1(j) \cdot \mathcal{B}_2(k) & i = -n \\ \sum_{j+k>n} \mathcal{B}_1(j) \cdot \mathcal{B}_2(k) & i = \infty \end{cases}$$

To ease readability we define $V(j,k,x,y) = \frac{\mathcal{B}_1(j)}{f^j} y(k) + \frac{\mathcal{B}_2(k)}{f^k} x(j) + x(j)y(k)$ *and based on* V *we define the EC terms as*

$$\tilde{\ell}_1 \times \tilde{\ell}_2(i) := \begin{cases} \sum_{j+k=i} V(j,k,\tilde{\ell}_1,\tilde{\ell}_2) & i \in N \setminus \{-n\} \\ \sum_{j+k \le -n} V(j,k,\tilde{\ell}_1,\tilde{\ell}_2) & i = -n \\ 0 & i = \infty \end{cases}$$

$$\ell_1 \times \ell_2(i) := \begin{cases} \sum_{j+k=i} V(j,k,\ell_1,\ell_2) & i \in N \setminus \{-n\} \\ 0 & i \in \{-n, \infty\} \end{cases}$$

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

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

where we define

$$\dagger\mathcal{B}_1(i) := \begin{cases} \mathcal{B}_1(2i-1) + \mathcal{B}_1(2i) & i \in [-n/2+1, n/2] \\ \mathcal{B}_1(\infty) & i = \infty \\ 0 & \textit{otherwise} \end{cases}$$

To ease the readability we define a function $W(i,x) := x(2i-1) + \mathcal{B}_1(2i-1)\left(\frac{1}{f_1^{2i-1}} - \frac{1}{f_1^{2i}}\right) + x(2i)$*. We define the EC terms as*

$$\dagger\tilde{\ell}_1(i) := \begin{cases} W(i,\tilde{\ell}_1) & i \in [-n/2+1, n/2] \\ \tilde{\ell}_1(-n) & i = -n/2 \\ 0 & \textit{otherwise} \end{cases}$$

$$\dagger\ell_1(i) := \begin{cases} W(i,\ell_1) & i \in [-n/2+1, n/2] \\ 0 & \textit{otherwise} \end{cases}$$

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

Definition 10 (Notation for composing ratio buckets)**.** *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 $(A_1, \ldots, A_j)$ and $(B_1, \ldots, B_j)$ and T_2 is over the distributions $(A_{j+1}, \ldots, A_{\mathcal{W}})$ and $(B_{j+1}, \ldots, B_{\mathcal{W}})$, we write

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

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

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

Whenever we say that ratio buckets $(\mathcal{B}, \tilde{\ell}, \ell, f, u)$ 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

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

where $(\text{🍃}(A_1, B_1), \ldots, \text{🍃}(A_{\mathcal{W}}, B_{\mathcal{W}}))$ 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 (EC) 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 EC 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 EC 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 11 (Ratio buckets with EC 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 $N = \{-n, \ldots, n\}$. We define the ratio buckets with EC terms per element as follows*

$$\begin{aligned}
\mathcal{B}(x) &:= P_A(x) \\
\tilde{\ell}(x) &:= \begin{cases} P_B(x) - \frac{P_A(x)}{f^{\iota_T(x)}} & \iota_{\text{🍃}(A,B)}(x) \in N, \\ 0 & \iota_{\text{🍃}(A,B)}(x) = \infty, \end{cases} \\
\ell(x) &:= \begin{cases} \tilde{\ell}(x) & \iota_{\text{🍃}(A,B)}(x) \in N \setminus \{-n\}, \\ 0 & \iota_{\text{🍃}(A,B)}(x) \in \{-n, \infty\}. \end{cases}
\end{aligned}$$

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 ι_T instead of the bucket indexes.

Definition 12 (Composition with EC terms per element)**.** *For ratio buckets $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ 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 EC terms as

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

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 EC terms as follows (where f is the old factor, from the label of T_1):

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

We now show our first important lemma for the soundness of our buckets and EC terms: the terms we defined just previously indeed characterize the impact of each individual event on the overall bucket values and EC 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 8 (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 ratio buckets (with EC 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 $i \in \{-n, \dots, n, \infty\}$:

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

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

- By definition, $\mathcal{B}(x) = P_A(x)$ (c.f., Definition 11). Thus, $\mathcal{B}(i) = \sum_{x \ s.t. \ \iota(x)=i} P_A(x) = \sum_{x \ s.t. \ \iota(x)=i} \mathcal{B}(x)$.
- If $i \in \{-n, \dots, n\}$, then $\tilde{\ell}(i) = \sum_{x \ s.t. \ \iota(x)=i} P_B(x) - \frac{P_A(x)}{f^i} = \sum_{x \ s.t. \ \iota(x)=i} \tilde{\ell}(x)$. Otherwise $\tilde{\ell}(i) = 0 = \sum_{x \ s.t. \ \iota(x)=i} 0 = \sum_{x \ s.t. \ \iota(x)=i} \tilde{\ell}(x)$.
- If $i \in \{-n+1, \dots, n\}$, then $\ell(i) = \sum_{x \ s.t. \ \iota(x)=i} P_B(x) - \frac{P_A(x)}{f^i} = \sum_{x \ s.t. \ \iota(x)=i} \ell(x)$. Otherwise $\ell(i) = 0 = \sum_{x \ s.t. \ \iota(x)=i} 0 = \sum_{x \ s.t. \ \iota(x)=i} \ell(x)$.

If $T = T_1 \times T_2$**:** We assume the lemma holds for ratio buckets $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ over a universe $\mathcal{U}_1$ and ratio buckets $(\mathcal{B}_2, \tilde{\ell}_2, \ell_2, f_2, u_2)$ 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 $i \in \{-n+1, \dots, n\}$

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

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

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

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 $\iota_T(x) = -n$ is equivalent to $\iota_{T_1}(x_1) + \iota_{T_2}(x_2) \leq -n$ and for ∞ we have $j + k > n$ and we know from Definition 7 that $\iota_T(x) = \infty$ is equivalent to $\iota_{T_1}(x_1) + \iota_{T_2}(x_2) \geq n$.

For the virtual error, we distinguish the following cases:

- $\iota_T(x) \in \{-n+1, \dots, n\}$. Then,

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

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

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

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

- $\iota_T(x) = \infty$.

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

If $T = \dagger T_1$:

We assume the lemma holds for ratio buckets $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ 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 $i \in \{-n, \dots, -n/2-1, n/2+1, \dots, n\}$

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

For $i = \infty$, we have

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

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

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

For $i = -n/2$ we have

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

We hence go forward to show the lemma for the EC terms.

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

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

For $i = \infty$, we have

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

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

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

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 $i \in \{-n+1, \ldots, n\}$ for $\ell(i)$.

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

With Lemma 8 we now have a powerful tool for proving a set of properties for our EC terms that will ultimately allow us to show the soundness of our results: We can relate every bucket value and every EC 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 EC 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 EC term is always smaller than the virtual EC term and finally we show that for every event x, the virtual EC term after an arbitrary amount of composition and squaring following the composition tree T still precisely captures $P_B(x) - \frac{P_A(x)}{f^{\iota_T}}$.

Lemma 9 (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 $k \in \{1, \ldots, \mathcal{W}\}$ $\mathcal{B}(A_k, B_k, f_k, n) = (\mathcal{B}_k, \tilde{\ell}_k, \ell_k, f_k, 1)$ be A_k/B_k ratio buckets (with EC terms) and let T be a composition tree. Let $\varepsilon \geq 0$. Let*

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

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

Proof. We show the lemma via structural induction over T. For leaf nodes $T = \varnothing(A, B)$, the real EC term of an initial bucketing is calculated as the sum of EC terms for each $x \in \mathcal{U}$, which are either $P_B(x) - \frac{P_A(x)}{f^{\iota_T(x)}}$ or 0. By definition we know that $P_A(x) \leq f^{\iota_T(x)} P_B(x)$, so all these values are positive. For composition $T_1 \times T_2$ we have either 0 or $V(j, k, x, y) = \frac{\mathcal{B}_1(j)}{f^j} y(k) + \frac{\mathcal{B}_2(k)}{f^k} x(j) + x(j)y(k)$, which is the sum and product of positive terms (the latter we know from the induction invariant). Analogously we notice that for squaring $\dagger T_1$ we have either 0 or $\ell_1(2i-1) + \mathcal{B}_1(2i-1)\left(\frac{1}{f^{2i-1}} - \frac{1}{f^{2i}}\right) + \ell_1(2i)$, which again consists purely of positive terms (again via induction invariant).

More precisely, we distinguish the following cases:

For $T = \varnothing(A, B)$, the real EC term of an initial bucketing is calculated as the sum of EC terms for each $x \in \mathcal{U}$ $\ell(x) = P_B(x) - \frac{P_A(x)}{f^{\iota_T(x)}}$ if $\iota_T(x) \notin \{n-, \infty\}$ and 0 otherwise. For $\iota_T(x) \in \{-n, \dots, n\}$ by definition we have $P_A(x) \leq f^{\iota_T(x)} P_B(x)$. 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 ratio buckets $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ and ratio buckets $(\mathcal{B}_2, \tilde{\ell}_2, \ell_2, f_2, u_2)$. By induction hypothesis, ℓ_1 and ℓ_2 are positive. We calculate the composed EC terms as either 0 (if $i \in \{-n, \infty\}$) or as

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

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

For $T = \dagger T_1$, We calculate the EC 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 EC terms and all bucket terms are positive. [5] Analogously, we can show that the virtual EC terms $\tilde{\ell}$ are positive as well. □

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

Lemma 10 (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 A_k/B_k ratio buckets (with EC 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 $T = \varnothing(A, B)$: 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, $\ell_T(x) \leq \tilde{\ell}_T(x)$.

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

For $T = T_1 \times T_2$**:** $\mathcal{B}_T$ is the result of composing ratio buckets $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ and ratio buckets $(\mathcal{B}_2, \tilde{\ell}_2, \ell_2, f_2, u_2)$. By induction hypothesis, $\ell_1 \leq \tilde{\ell}_1$ and $\ell_2 \leq \tilde{\ell}_2$. For $\iota_T(x) = -n$, $\ell(x) = 0$. By Lemma 9 we know that $0 \leq \tilde{\ell}(x)$, hence $\ell(x) = 0 \leq \tilde{\ell}(x)$. For $\iota_T(x) \neq -n$, with $x_1 \in \mathcal{U}_1$ and $x_2 \in \mathcal{U}_2$ we have

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

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 EC 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 EC term. For our upper bound on δ we circumvent this problem by over-approximating misplaced events (using the real EC term) and by not using EC terms in some buckets with a bucket factor f^i close to e^ε.

Lemma 11 (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 $k \in \{1, \ldots, \mathcal{W}\}$ $\mathcal{B}(A_k, B_k, f_k, n) = (\mathcal{B}_k, \tilde{\ell}_k, \ell_k, f_k, 1)$ be A_k/B_k ratio buckets (with EC terms) and let T be a composition tree. Let $\varepsilon \geq 0$,*

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

Then, for $x \in \mathcal{U}$ with $\iota_T(x) \neq \infty$ we have

$$\tilde{\ell}(x) = P_B(x) - \frac{P_A(x)}{f^{\iota_T(x)}}$$

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

$$P_B(x) - \frac{P_A(x)}{f^i} = P_{B_1}(x) - \frac{P_{A_1}(x)}{f^i} = \tilde{\ell}(i),$$

where f is the bucketing factor of the leaf.

For $T = T_1 \times T_2$, $\mathcal{B}_T$ is the result of composing ratio buckets $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ and ratio buckets $(\mathcal{B}_2, \tilde{\ell}_2, \ell_2, f_2, u_2)$, 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 EC term composition we get with

$x_1 \in \mathcal{U}_1$ and $x_2 \in \mathcal{U}_2$

$$
\begin{aligned}
\tilde{\ell}(x) =& (\tilde{\ell}_1 \times \tilde{\ell}_2)(x) \\
=& \left(\frac{P_{A_1}(x_1)}{f^{\iota_{T_1}(x_1)}} + \tilde{\ell}_1(x_1)\right)\tilde{\ell}_2(x_2) + \left(\frac{P_{A_2}(x_2)}{f^{\iota_{T_2}(x_2)}} + \tilde{\ell}_2(x_2)\right)\tilde{\ell}_1(x_1) - \tilde{\ell}_1(x_1)\tilde{\ell}_2(x_2) \\
=& \frac{P_{A_1}(x_1)}{f^{\iota_{T_1}(x_1)}} \cdot \tilde{\ell}_2(x_2) + \frac{P_{A_2}(x_2)}{f^{\iota_{T_2}(x_2)}} \cdot \tilde{\ell}_1(x_1) + \tilde{\ell}_1(x_1)\tilde{\ell}_2(x_2) \\
\overset{IH}{=}& \frac{P_{A_1}(x_1)}{f^{\iota_{T_1}(x_1)}} \cdot \left(P_{B_2}(x_2) - \frac{P_{A_2}(x_2)}{f^{\iota_{T_2}(x_2)}}\right) \\
&+ \frac{P_{A_2}(x_2)}{f^{\iota_{T_2}(x_2)}} \cdot \left(P_{B_1}(x_1) - \frac{P_{A_1}(x_1)}{f^{\iota_{T_1}(x_1)}}\right) \\
&+ \left(P_{B_1}(x_1) - \frac{P_{A_1}(x_1)}{f^{\iota_{T_1}(x_1)}}\right) \cdot \left(P_{B_2}(x_2) - \frac{P_{A_2}(x_2)}{f^{\iota_{T_2}(x_2)}}\right) \\
=& \frac{P_{A_1}(x_1)}{f^{\iota_{T_1}(x_1)}} \cdot P_{B_2}(x_2) - \frac{P_A(x)}{f^{\iota_T(x)}} \\
&+ \frac{P_{A_2}(x_2)}{f^{\iota_{T_2}(x_2)}} \cdot P_{B_1}(x_1) - \frac{P_A(x)}{f^{\iota_T(x)}} \\
&+ P_B(x) - \frac{P_{A_1}(x_1)}{f^{\iota_{T_1}(x_1)}} \cdot P_{B_2}(x_2) - \frac{P_{A_2}(x_2)}{f^{\iota_{T_2}(x_2)}} \cdot P_{B_1}(x_1) + \frac{P_A(x)}{f^{\iota_T(x)}} \\
=& P_B(x) - \frac{P_A(x)}{f^{\iota_T(x)}}
\end{aligned}
$$

For $T = \dagger T_1$, we know that for all $x \in \mathcal{U}$, $\iota_{T_1}(x) \in \{-n/2, \ldots, n/2\} \cup \{\infty\}$. 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

$$\tilde{\ell}_1(x) = P_B(x) - \frac{P_{A_1}(x)}{f^{\iota_T(x)}}$$

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

$$
\begin{aligned}
\tilde{\ell}(x) = \dagger\tilde{\ell}_1(x) = \tilde{\ell}_1(x) + \mathcal{B}_{A_1}(x) \cdot \left(\frac{1}{f^{\iota_{T_1}(x)}} - \frac{1}{f^{2\cdot\lceil \iota_{T_1}(x)/2 \rceil}}\right) \\
\overset{IH}{=} P_B(x) - \frac{P_A(x)}{f^{\iota_{T_1}(x)}} + P_A(x) \cdot \left(\frac{1}{f^{\iota_{T_1}(x)}} - \frac{1}{f^{2\cdot\lceil \iota_{T_1}(x)/2 \rceil}}\right) \\
= P_B(x) - \frac{P_A(x)}{f^{\iota_{T_1}(x)}} + P_A(x) \cdot \left(\frac{1}{f^{\iota_{T_1}(x)}} - \frac{1}{f^{2\iota_T(x)}}\right) \\
= P_B(x) - \frac{P_A(x)}{(f^2)^{\iota_T(x)}}.
\end{aligned}
$$

□

4.4 The approximated delta with error correction

Finally, we define how to calculate a sound upper bound on δ based on ratio buckets with EC terms. We note that when using the real EC 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 events into the next larger bucket. This slight misplacement poses a problem for a small number of buckets with a bucket factor f^i just slightly larger than e^ε, as they can now contain events that should have been placed in a lower bucket (with factor $f^{i^*} < e^{\varepsilon}$) and that now actually have a negative contribution to δ: $P_A(x) - e^{\varepsilon} P_B(x) < 0$. All sets of A/B ratio buckets carry 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 EC term for buckets with index $i < j_\varepsilon + u$and instead fall back to Definition 4 for those buckets. For the remaining buckets with $i \geq j_\varepsilon + u$, which typically is the vast majority of buckets, we make use of the real EC term to reduce the error.

Definition 13 (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* A_k/B_k *ratio buckets (with EC 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*

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

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

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

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

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

Lemma 12 (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 ratio buckets (with EC 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 $j_\varepsilon + u \leq \iota_T(x) \neq \infty$*, then the EC term never makes a negative contribution to the approximated delta with EC:* $\ell(x) \leq \frac{P_A(x)}{f^{\iota_T(x) - u}} - \frac{P_A(x)}{f^{\iota_T(x)}}$*.*

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

Let $T =$ 🍃(A, B)**.** If $\iota_T(x) = -n$ then

$$\ell(x) = 0 \leq P_A(x) \cdot \left(\frac{1}{f^{-n-1}} - \frac{1}{f^{-n}} \right).$$

Otherwise, if $\iota_T(x) > -n$, we know that by definition of $\iota_T(x)$ we have $f^{\iota_T(x) - 1} P_B(x) \leq P_A(x)$

$$\begin{aligned} \ell(x) =& P_B(x) - \frac{P_A(x)}{f^{\iota_T(x)}} \\ \leq& \frac{P_A(x)}{f^{\iota_T(x) - 1}} - \frac{P_A(x)}{f^{\iota_T(x)}}. \end{aligned}$$

Let $T = T_1 \times T_2$. If $\iota_T(x) = -n$, then $\ell(x) = 0 \leq \frac{P_A(x)}{f^{\iota_T(x)-u}} - \frac{P_A(x)}{f^{\iota_T(x)}}$. Otherwise, $\mathcal{B}_T$ is the result of composing ratio buckets $(\mathcal{B}_1, \tilde{\ell}_1, \ell_1, f_1, u_1)$ and ratio buckets $(\mathcal{B}_2, \tilde{\ell}_2, \ell_2, f_2, u_2)$. 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 $\iota_T(x) = \iota_{T_1}(x_1) + \iota_{T_2}(x_2)$. Moreover, we know that $P_A(x) = P_{A_1}(x_1) \cdot P_{A_2}(x_2)$. Thus, for $u = u_1 + u_2$ we get

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

$$\begin{aligned}
&\overset{IH}{\leq} \left(\frac{P_{A_1}(x_1)}{f^{\iota_{T_1}(x_1)}}\right)\left(\frac{P_{A_2}(x_2)}{f^{\iota_{T_2}(x_2)-(u-u_1)}} - \frac{P_{A_2}(x_2)}{f^{\iota_{T_2}(x_2)}}\right) \\
&\quad + \left(\frac{P_{A_2}(x_2)}{f^{\iota_{T_2}(x_2)}}\right)\left(\frac{P_{A_1}(x_1)}{f^{\iota_{T_1}(x_1)-u_1}} - \frac{P_{A_1}(x_1)}{f^{\iota_{T_1}(x_1)}}\right) \\
&\quad + \left(\frac{P_{A_1}(x_1)}{f^{\iota_{T_1}(x_1)-u_1}} - \frac{P_{A_1}(x_1)}{f^{\iota_{T_1}(x_1)}}\right)\left(\frac{P_{A_2}(x_2)}{f^{\iota_{T_2}(x_2)-(u-u_1)}} - \frac{P_{A_2}(x_2)}{f^{\iota_{T_2}(x_2)}}\right) \\
&= \frac{P_A(x)}{f^{\iota_T(x)-(u-u_1)}} - \frac{P_A(x)}{f^{\iota_T(x)}} \\
&\quad + \frac{P_A(x)}{f^{\iota_{T_2}(x_2)+\iota_{T_1}(x_1)-u_1}} - \frac{P_A(x)}{f^{\iota_{T_2}(x_2)+\iota_{T_1}(x_1)}} \\
&\quad + \frac{P_A(x)}{f^{\iota_T(x)-u_1-(u-u_1)}} - \frac{P_A(x)}{f^{\iota_T(x)-(u-u_1)}} - \frac{P_A(x)}{f^{\iota_T(x)-u_1}} + \frac{P_A(x)}{f^{\iota_T(x)}} \\
&= \frac{P_A(x)}{f^{\iota_T(x)-u_2}} - \frac{P_A(x)}{f^{\iota_T(x)}} \\
&\quad + \frac{P_A(x)}{f^{\iota_T(x)-u_1}} - \frac{P_A(x)}{f^{\iota_T(x)}} \\
&\quad + \frac{P_A(x)}{f^{\iota_T(x)-u}} - \frac{P_A(x)}{f^{\iota_T(x)-u_2}} - \frac{P_A(x)}{f^{\iota_T(x)-u_1}} + \frac{P_A(x)}{f^{\iota_T(x)}} \\
&= \frac{P_A(x)}{f^{\iota_T(x)-u}} - \frac{P_A(x)}{f^{\iota_T(x)}}
\end{aligned}$$

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 $\dagger\ell(x) = \ell_1(x) + \mathcal{B}_1(x) \cdot \left(\frac{1}{f_1^{\iota_{T_1}(x)}} - \frac{1}{f_1^{2 \cdot \lceil \iota_{T_1}(x)/2 \rceil}}\right)$. Since we excluded $\iota_T = \infty = \iota_{T_1}$ and $j_\varepsilon + u \leq \iota_T$, we know that $\iota_T \in \{0, \ldots, n/2\}$.

Thus,

$$\begin{aligned}
\ell(x) &= \dagger\ell_1(x) \\
&= \ell_1(x) + \mathcal{B}_1(x) \cdot \left(\frac{1}{f_1^{\iota_{T_1}(x)}} - \frac{1}{f_1^{2\cdot\lceil \iota_{T_1}(x)/2 \rceil}} \right) \\
&\overset{IH}{\leq} \frac{P_A(x)}{f_1^{\iota_{T_1}(x)-u_1}} - \frac{P_A(x)}{f_1^{\iota_{T_1}(x)}} + \mathcal{B}_1(x) \cdot \left(\frac{1}{f_1^{\iota_{T_1}(x)}} - \frac{1}{f_1^{2\cdot\lceil \iota_{T_1}(x)/2 \rceil}} \right) \\
&= \frac{P_A(x)}{f_1^{\iota_{T_1}(x)-u_1}} - \frac{P_A(x)}{f_1^{\iota_{T_1}(x)}} + \frac{P_A(x)}{f_1^{\iota_{T_1}(x)}} - \frac{P_A(x)}{f_1^{2\cdot\lceil \iota_{T_1}(x)/2 \rceil}} \\
&= \frac{P_A(x)}{(f_1^2)^{\frac{\iota_{T_1}(x)-u_1}{2}}} - \frac{P_A(x)}{(f_1^2)^{\iota_T(x)}} \\
&\leq \frac{P_A(x)}{(f_1^2)^{\lceil \iota_{T_1}(x)/2 \rceil - (\lceil u_1/2 \rceil + 1)}} - \frac{P_A(x)}{(f_1^2)^{\iota_T(x)}} \\
&= \frac{P_A(x)}{(f_1^2)^{\iota_T(x)-u}} - \frac{P_A(x)}{(f_1^2)^{\iota_T(x)}}
\end{aligned}$$

□

From Lemma 12 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 13 (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 A_k/B_k ratio buckets (with EC 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:*

$$\delta(\mathcal{B}_T, \varepsilon) \geq \sum_{x\in\mathcal{U}} \max\left(0, P_A(x) - e^\varepsilon P_B(x)\right)$$

Proof. As $\delta(\mathcal{B}_T, \varepsilon)$ is is a sum over $\mathcal{B}_A$ and ℓ, Lemma 8 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 $j_\varepsilon \leq \iota_T(x) \leq j_\varepsilon + u - 1$. By Lemma 6, we know the following

$$\begin{aligned}
P_A(x) \leq& f^{\iota_T(x)} P_B(x) \\
\Leftrightarrow \frac{P_A(x)}{f^{\iota_T(x)}} \leq& P_B(x) \\
\Leftrightarrow P_A(x) - e^\varepsilon \frac{P_A(x)}{f^{\iota_T(x)}} \geq& P_A(x) - e^\varepsilon P_B(x)
\end{aligned}$$

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

$$\delta(\mathcal{B}_T, x, \varepsilon) = P_A(x) - e^\varepsilon \frac{P_A(x)}{f^{\iota_T(x)}} \geq P_A(x) - e^\varepsilon P_B(x)$$

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

$$\delta(\mathcal{B}_T, x, \varepsilon) = P_A(x) - \underbrace{\frac{e^\varepsilon}{f^{\iota_T(x)}}}_{\le 1} P_A(x) \ge 0$$

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

We show two things: (i)

$$P_A(x) - e^\varepsilon \left(\frac{P_A(x)}{f^{\iota_T(x)}} + \ell(x) \right)$$

by Lemma 12 we know that $\ell(x) \le \frac{P_A(x)}{f^{\iota_T(x)-u}} - \frac{P_A(x)}{f^{\iota_T(x)}}$ holds; hence, we get

$$\begin{aligned}
&\ge P_A(x) - e^\varepsilon \left(\frac{P_A(x)}{f^{\iota_T(x)}} + \frac{P_A(x)}{f^{\iota_T(x)-u}} - \frac{P_A(x)}{f^{\iota_T(x)}} \right) \\
&= P_A(x) - e^\varepsilon \left(\frac{P_A(x)}{f^{\iota_T(x)-u}} \right) \\
&\ge P_A(x) - \frac{f^{j_\varepsilon}}{f^{\iota_T(x)-u}} P_A(x) \\
&= P_A(x) \cdot \left(1 - \frac{f^{j_\varepsilon}}{f^{\iota_T(x)-u}} \right)
\end{aligned}$$

as by assumption $\iota_T(x) \ge j_\varepsilon + u$, we get

$$\ge 0$$

(ii) Note that

$$\frac{P_A(x)}{f^{\iota_T(x)}} + \underbrace{\ell(x)}_{\le \tilde{\ell}(x)} \le \frac{P_A(x)}{f^{\iota_T(x)}} + \tilde{\ell}(x) \overset{Lemma\ 11}{=} \frac{P_A(x)}{f^{\iota_T(x)}} + P_B(x) - \frac{P_A(x)}{f^{\iota_T(x)}} = P_B(x)$$

Thus,

$$P_A(x) - e^\varepsilon \left(\frac{P_A(x)}{f^{\iota_T(x)}} + \ell(x) \right) \ge P_A(x) - e^\varepsilon P_B(x)$$

From (i) and (ii) we get

$$\begin{aligned}
\delta(A, B, x, f, n, u, \varepsilon) =& P_A(x) - e^\varepsilon \left(\frac{P_A(x)}{f^{\iota_T(x)}} + \ell(x) \right) \\
&\ge \max(0, P_A(x) - e^\varepsilon P_B(x))
\end{aligned}$$

Case 3. By definition of δ, we have $\delta(x) = P_A(x) > \max\left(0, P_A(x) - e^\varepsilon P_B(x)\right)$.

Case 4. Thus, for all x with $\iota_T(x) \le j_\varepsilon$,

$$\begin{aligned}
& P_A(x) - e^\varepsilon P_B(x) \\
\le & P_A(x) - f^{j_\varepsilon} P_B(x) \\
\le & P_A(x) - f^{\iota_T(x)} P_B(x) \overset{\text{Lemma 6}}{\le} 0
\end{aligned}$$

and thus,

$$\delta(x) = 0 = \max\left(0, P_A(x) - e^\varepsilon P_B(x)\right)$$

□

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 13 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 EC 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 A_k/B_k ratio buckets (with EC terms) and let T be a composition tree. Let $\varepsilon \geq 0$ and $j_\varepsilon \in \mathbb{N}$ s.t. $f^{j_\varepsilon - 1} < e^\varepsilon \leq f^{j_\varepsilon}$,*

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

$$\delta_\varepsilon = \max\left(\sum_{x \in \mathcal{U}} \max\left(P_A(x) - e^\varepsilon P_B(x), 0\right),\right.$$

$$\left.\sum_{x \in \mathcal{U}} \max\left(P_B(x) - e^\varepsilon P_A(x), 0\right)\right)$$

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

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 13 proves that $\delta_\varepsilon \leq \delta(\mathcal{B}_T, \varepsilon)$ holds true.

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

$$\begin{aligned}
\delta^{\text{low}} &= \sum_{i \in \{j_\varepsilon, \ldots, n\}} \max\left(0, \mathcal{B}(i) - e^\varepsilon\left(\frac{\mathcal{B}(i)}{f^i} + \tilde{\ell}(i)\right)\right) \\
&\overset{\text{Lemma 8}}{=} \sum_{i \in \{j_\varepsilon, \ldots, n\}} \max\left(0, \sum_{x \in \mathcal{U}, \iota_T(x) = i} P_A(x) - e^\varepsilon\left(\frac{P_A(x)}{f^i} + \tilde{\ell}(x)\right)\right) \\
&\overset{\text{Lemma 11}}{=} \sum_{i \in \{-n, \ldots, n, \infty\}} \max\left(0, \sum_{x \in \mathcal{U}, \iota_T(x) = i} P_A(x) - e^\varepsilon\left(\frac{P_A(x)}{f^i} + \left(P_B(x) - \frac{P_A(x)}{f^{\iota_T(x)}}\right)\right)\right) \\
&= \sum_{i \in \{j_\varepsilon, \ldots, n\}} \max\left(0, \sum_{x \in \mathcal{U}, \iota_T(x) = i} P_A(x) - e^\varepsilon P_B(x)\right) \\
&\leq \sum_{i \in \{j_\varepsilon, \ldots, n\}} \sum_{x \in \mathcal{U}, \iota_T(x) = i} \max\left(0, P_A(x) - e^\varepsilon P_B(x)\right) \\
&\leq \sum_{x \in \mathcal{U}} \max\left(0, P_A(x) - e^\varepsilon P_B(x)\right)
\end{aligned}$$

Hence, we conclude that

$$\delta^{\text{low}} \leq \sum_{x \in \mathcal{U}} \max\left(0, P_A(x) - e^\varepsilon P_B(x)\right) = \delta_\varepsilon$$

□

Thus, the bounds calculated present a sound over-approximation of the real differential privacy values. We emphasize that distributions can be used to calculate differential privacy in a variety of applications. We make the notion of worst-case inputs formal. We require the existence of *worst-case* inputs that allow us to directly derive the relevant distributions.

Definition 14 (Worst-case inputs). *Inputs x_0, x_1 are worst-case inputs for a given sensitivity s and a mechanism M if $\Pr[M(x_0) \in S] \leq e^{\varepsilon} \Pr[M(x_1) \in S] + \delta$, implies M is (ε, δ)-ADP for all inputs with sensitivity s.*

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

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, for instance, whenever worst-case inputs 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 (see Section 1.3).

4.5 Implementation

We implemented $\delta(\mathcal{B}_T, \varepsilon)$ (c.f. Theorem 2) in Python using the NumPy [2] and the SciPy [3] libraries in 655 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 efficient than our current implementation.

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

Given any ratio buckets 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 compositions 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 13 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 [12] that significantly improves on the standard and advanced composition theorem. This composition theorem [12] provides a composition result where each ε, δ pair after r-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 ratio buckets 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 ratio buckets can be seen as an approximation of the two ε, δ graphs[6] 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 [16]. While our approach does not provide significantly tighter bounds for Laplace

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

mechanism, our ratio buckets significantly improve 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 ratio buckets 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.

5.1 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 $\mathrm{LP}(\mu, \gamma)$. As a result, an attacker either makes observations from $\mathrm{LP}(0, 200)$ or from $\mathrm{LP}(1, 200)$ (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 ratio buckets 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 bucket method introduced in Definition 8 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 x such that $f^i < p_A(x)/p_B(x) \leq f^{i+1}$. 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 $\mathrm{Laplace}(x) := \frac{1}{2\gamma} e^{\frac{-|x-\mu|}{\gamma}}$. 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 $(\varepsilon, 0)$-ADP with $\varepsilon = \frac{1}{\gamma}$. Consequently, we can describe the quotient f at each point x as We calculate the quotient $f(x) = \frac{\mathrm{Laplace}_{\mu_1}(x)}{\mathrm{Laplace}_{\mu_2}(x)}$ 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* $\mathrm{lbb}(i)$ and the value x with $f(x) = f^i$ the *right bucket border* $\mathrm{rbb}(i)$.

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 $\mathrm{LP}(0, 200)$), denoted as $\mathrm{GS}(\mu, \sigma^2)$, 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 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

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. [16] 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.[7]

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 Kairouz et al.'s composition theorem

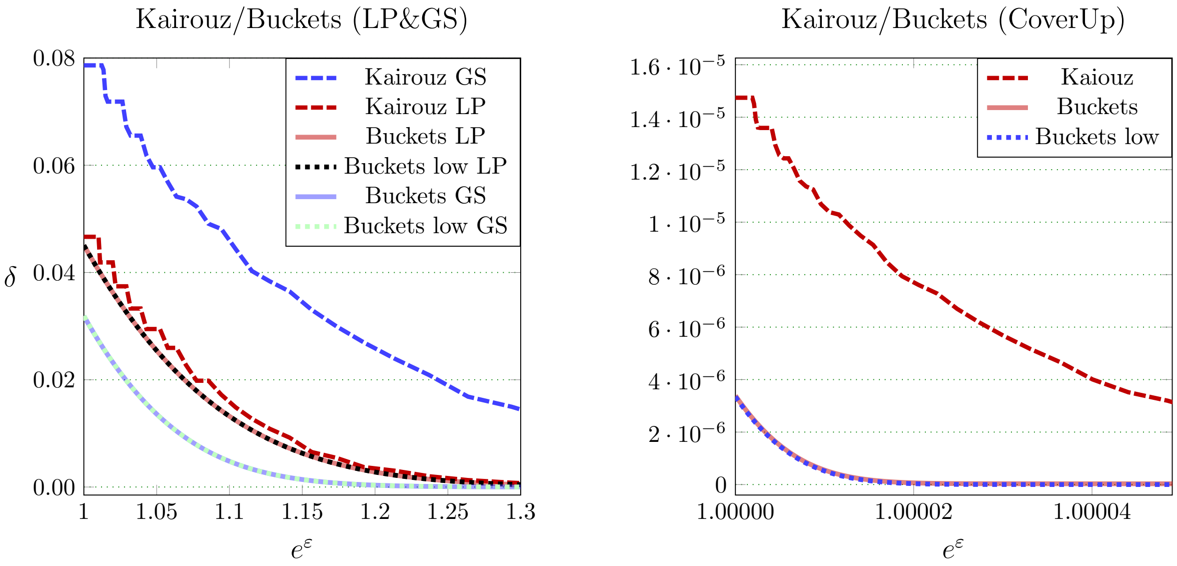
We directly implement the bounds from Kairouz et al.'s theorem. We do not use any statements specific to Gauss or Laplace, as those are simplified and provide worse bounds.

Theorem 3 ([12])**.** *For any $\varepsilon \geq 0$ and $\delta \in [0, 1]$, the class of (ε, δ)-ADP mechanisms satisfies (ε', δ')-ADP under r-fold composition, for all $i \in \{0, \ldots, \lfloor r/2 \rfloor\}$ where $\varepsilon' = (r - 2i)\varepsilon$ and $\delta' = 1 - (1 - \delta)^r (1 - \delta_i)$*

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

We compute for a given number r of compositions the epsilon-delta graph 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 $i \in \{0, \ldots, \lfloor r/2 \rfloor\}$. From these stored (ε', δ') pairs,

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



(a) Laplace LP$(5k, 200)$ and Gauss GS$(5k, 2 \cdot 200^2)$.

(b) CoverUp data.

Figure 8: The ε, δ graphs computed with Kairouz et al.'s composition theorem and with our ratio buckets after 512 compositions for the Laplace mechanism, the Gaussian mechanism, and the CoverUp data (upper bound solid, lower bound dotted). 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.

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 $r = 512$ 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

$$e^\varepsilon \in \{(1 + 10^{-14})^{1.1^j} \mid j \in \{0, \ldots, n\}\},$$

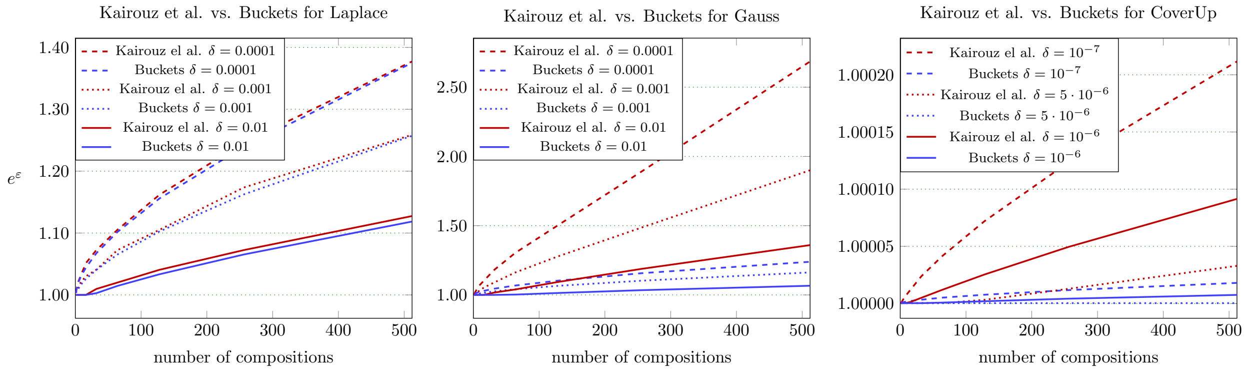
where we choose n as a point where the (ε, δ) after r-fold composition becomes stationary. While we concede that it might be possible to obtain a slightly lower bound from the composition theorem, we are confident that, due to this fine grid, the resulting graphs for Kairouz et al.'s composition theorem that we compute are representative.

5.5 Comparing evaluations

We are finally in a position to evaluate how our ratio buckets compare against Kairouz et al.'s composition theorem. Figure 8a 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 8a shows that for the Gaussian mechanism that composition theorem is already after 512 compositions not very tight. Figure 8b shows that for the CoverUp-data our ratio buckets are tight, while there is a large gap to the bounds from Kairouz et al's composition theorem.

Figure 9 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.

The tightness of Kairouz et al.'s[12] bounds for the Laplace mechanism suggests that there is no noise distribution with the same, or smaller, initial ε and δ values that has a worse composition behavior than the



(a) Laplace

(b) Gauss

(c) CoverUp

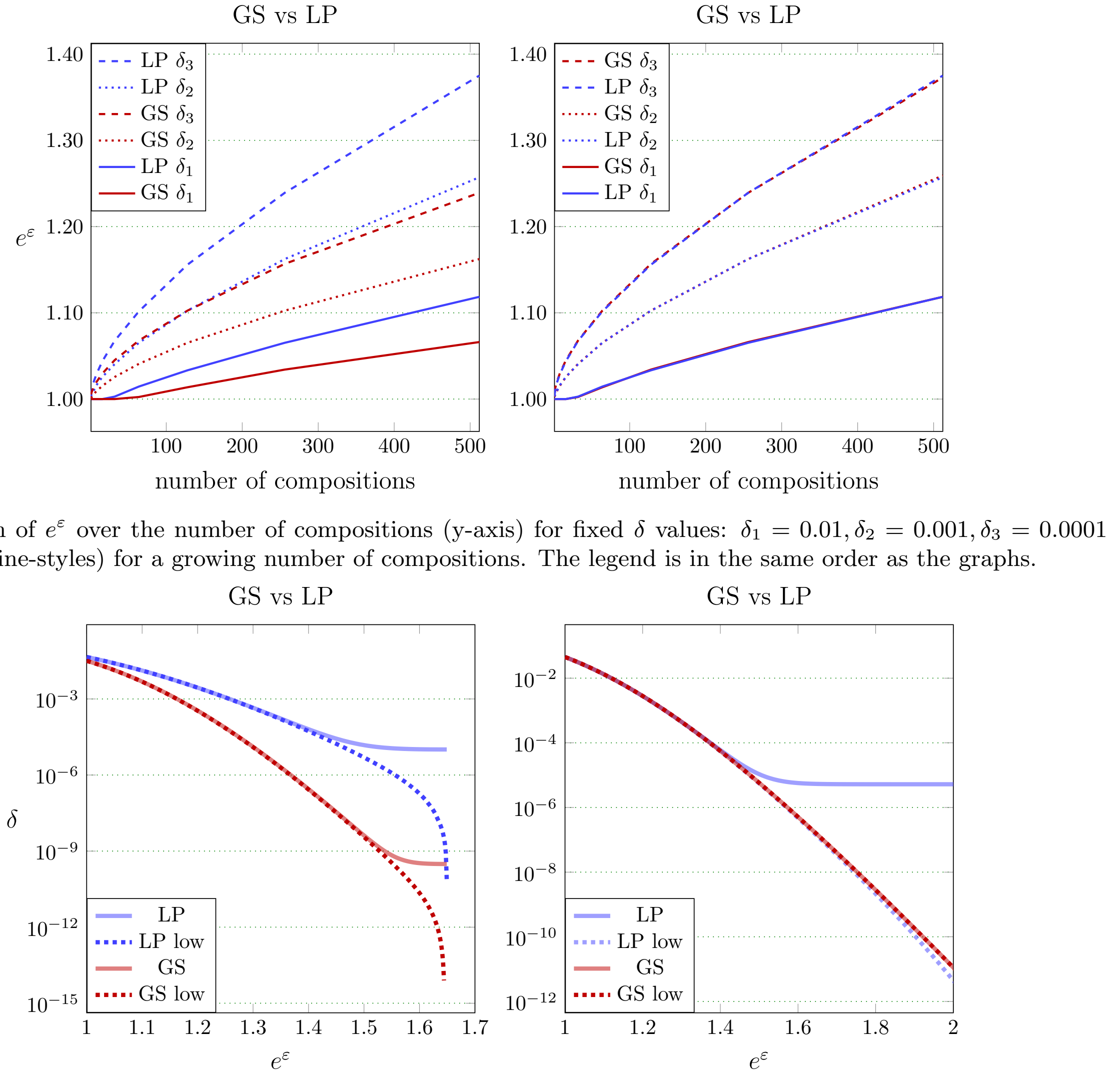
Figure 9: Growth of e^{ε} over the number of compositions (y-axis) for fixed δ values (different line-styles) for a growing number of compositions with mechanisms as in Figure 8.

Laplace mechanism.

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 10 (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 10 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 lower left graph in Figure 10 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.

Additionally, we found evidence that the epsilon-delta graph of the Laplace mechanism converges toward the epsilon-delta graph of a Gaussian mechanism with half the variance of the Laplace mechanism. For the same sensitivity, and truncations as above, the two right graphs in Figure 10 illustrate that after 512 compositions these two graphs converge toward each other. The upper right graph in Figure 10 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 10 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.



(a) Growth of e^ε over the number of compositions (y-axis) for fixed δ values: $\delta_1 = 0.01, \delta_2 = 0.001, \delta_3 = 0.0001$ (different line-styles) for a growing number of compositions. The legend is in the same order as the graphs.

(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 10: 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.

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. As a case study, we use the Vuvuzela [18] protocol, 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 not only estimate tighter bounds for the Laplace noise but also propose to change the shape of the noise distribution to Gaussian noise. With our bucketing approach, we show that already 5 to 10 times less noise[8] suffices to achieve the same strong privacy properties. [9]

[8] 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.

[9] We acknowledge that for the analysis of the Laplace noise previous results [12] would already yield tight results, but for the Gaussian noise our approach yields much tighter results (see Section 6).

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.

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.

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 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 $\lceil \max(0, \text{Laplace}(\mu_c/2, \gamma_c/2)) \rceil$. The system parameters $\mu_d, \mu_c, \gamma_d, \gamma_c$ 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 $\max(0, \text{Laplace}(\mu_d, \gamma_d))$ and $\max(0, \text{Laplace}(\mu_d + 2, \gamma_d))$, i.e., the (ϵ, δ) differential privacy of the product distributions $\max(0, \text{Laplace}(\mu_d, \gamma_d))^r$ and $\max(0, \text{Laplace}(\mu_d + 2, \gamma_d))^r$. 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 $\max(0, \text{Laplace}(\mu_c, \gamma_c)) + \max(0, \text{Laplace}(\mu_c/2, \gamma_c/2))$ and $\max(0, \text{Laplace}(\mu_c + 2, \gamma_c)) + \max(0, \text{Laplace}(\mu_c/2 + 1, \gamma_c/2))$. 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.

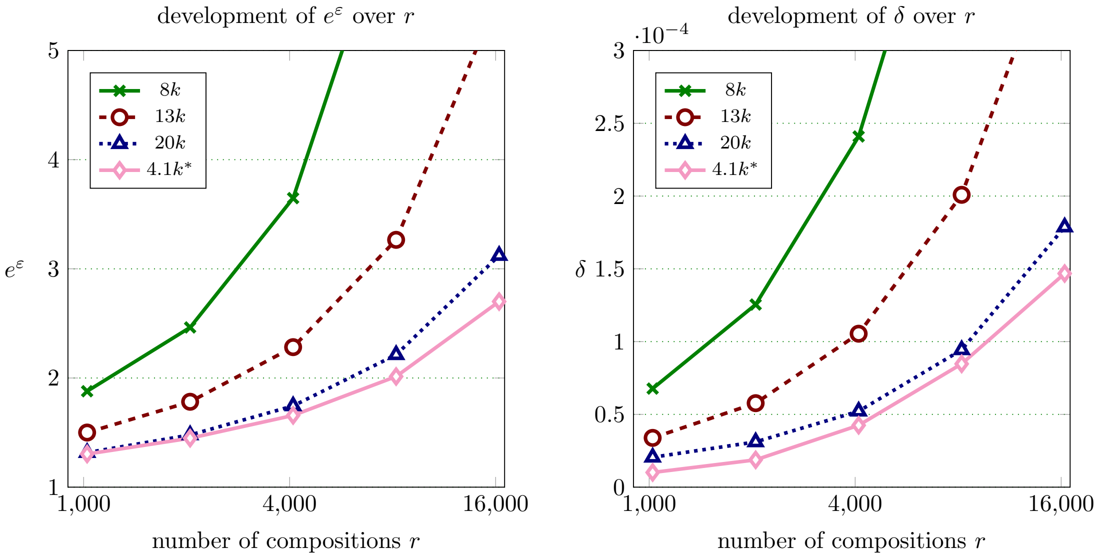


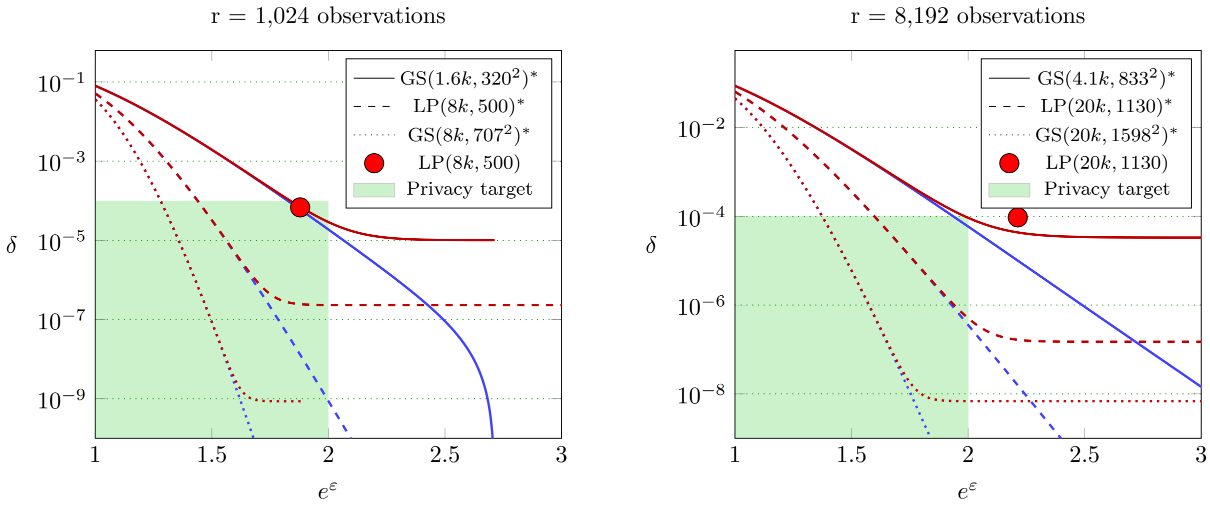
Figure 11: The privacy bounds for Vuvuzela's dialing protocol. The left graph shows the e^ε-values on the y-axis and the number of observations r on the x-axis (i.e., r-fold composition) in log-scale and the right graph shows the corresponding δ-values on the y-axis. The solid green ($\mu = 8k, \gamma = 500$), 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.

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 noise 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.

7.2 Tighter privacy analysis for the dialing protocol

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. We refer to Figures 13 and 16, placed in the appendix. 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 11 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 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 $\mathrm{LP}(20k, 1130)$ from the original Vuvuzela paper. The figure shows that our bounds with the reduced noise and with using Gaussian noise $\mathrm{GS}(4.1k, 833^2)$ are below the previous bounds for the highest noise configuration $\mathrm{LP}(20k, 1130)$, 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



(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 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, $\delta \leq 10^{-4}, e^{\varepsilon} \leq 2$).

$\delta \leq 10^{-4}$.

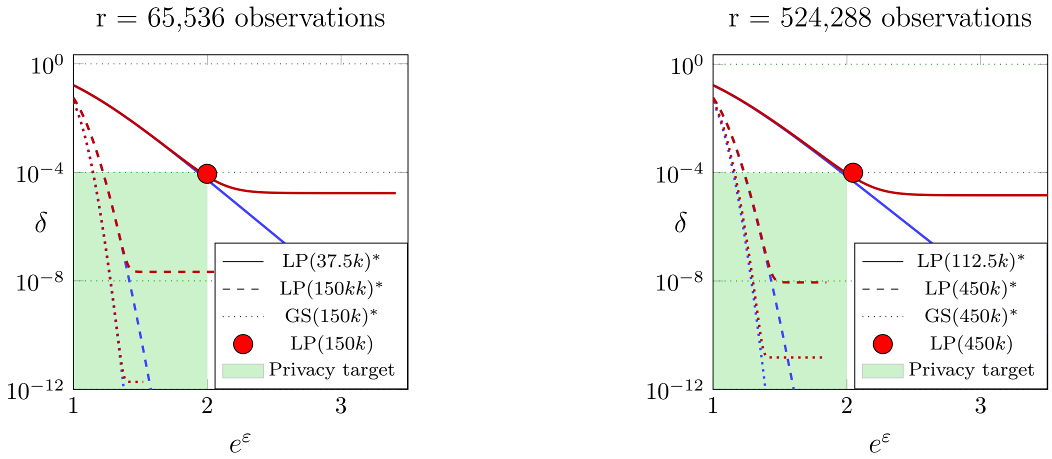
Next, we illustrate that our method computes bounds that are several orders of magnitude better than Vuvuzela's original bounds. For $r = 8,192$ observations, Figure 12b illustrates that using the highest noise configuration with Laplace noise LP($20k, 1130$) results in a privacy bound that is almost 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.[10]

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 12a 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 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 16 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

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



(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 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 conversation protocol). The figure depicts upper (red) and a lower bounds (blue) and Vuvuzela's privacy target (green, $\delta \leq 10^{-4}, e^\varepsilon \leq 2$).

noise can have in practice.

7.3 Tighter privacy analysis for the conversation protocol

Figure 14 depicts the epsilon graph and the delta graph with increasing r, respectively, for the conversation protocol. We compare Gaussian noise GS-new$_2$ with the previous bounds for the recommended noise configurations. The figure shows 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 $\delta \leq 10^{-4}$.

For $r = 524,288$ observations, Figure 15b shows that using LP-high results in bounds for δ that are almost 4 orders of magnitude lower, and for the corresponding Gaussian noise GS-high more than 6 orders of magnitude in comparison to their original result. Also, Figure 15b shows the corresponding lower bounds. We can see that our bounds for the reduced noise configuration GS-new$_2$ are tight up to $\delta \geq 10^{-5}$, for LP-high up to $\delta \geq 10^{-8}$, and for GS-high up to $\delta \approx 10^{-10}$ for reasonably small values of ε. Furthermore, we can see that GS-high 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 15a the ε, δ graphs for the conversation protocol for $r = 65,536$. Here, we can also see the tightness of our bound: for LP-low up to $\delta \geq 10^{-7}$, for GS-low up to $\delta \geq 10^{-11}$, and for GS-new$_1$ up to $\delta \geq 10^{-6}$. We can see that GS-low is more than 7 orders of magnitude lower than Vuvuzela's previous bounds for the same noise level. Moreover, we can see that GS-low meets and even 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 13 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 *ratio buckets*, a sound numerical approach for computing upper and lower bounds for differential privacy after r-fold composition. Our approach is based on concrete distributions,

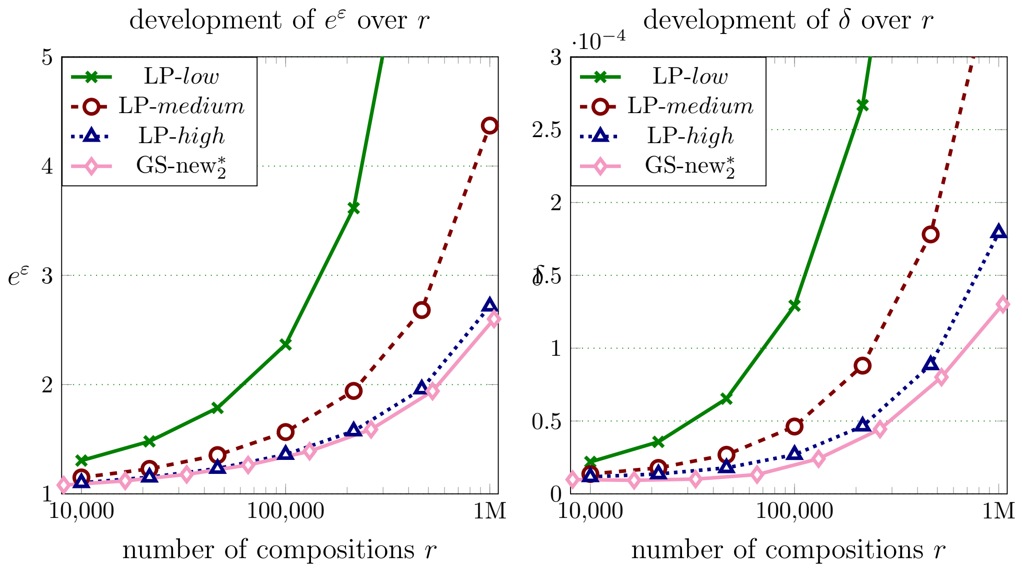
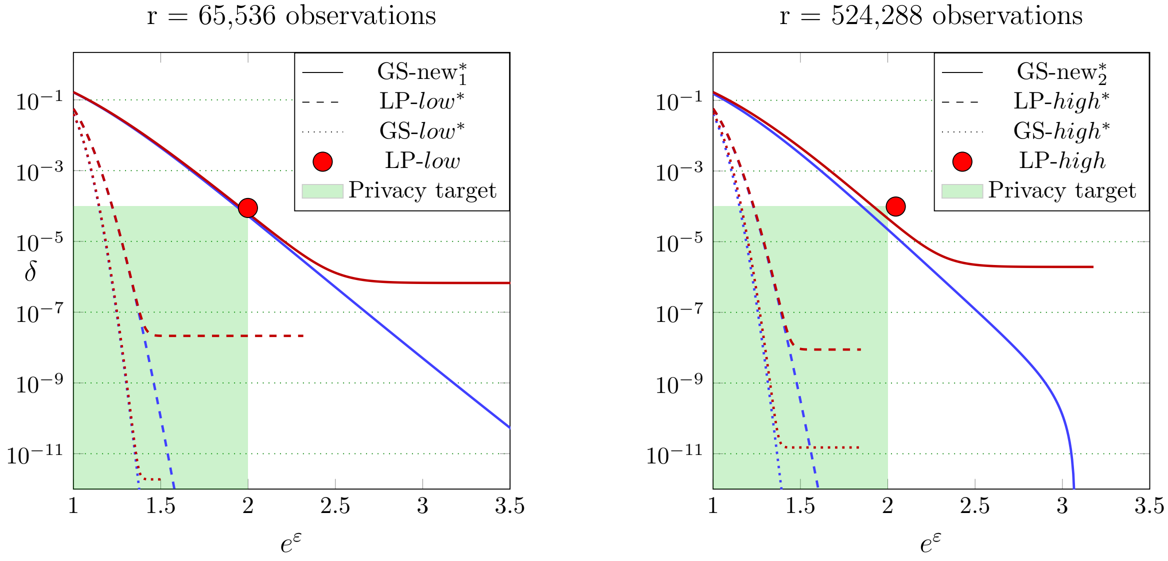


Figure 14: The privacy bounds for Vuvuzela's conversion protocol. The left graph shows the e^ε-values on the y-axis and the number of observations r on the x-axis (i.e., r-fold composition) in log-scale and the right graph shows the corresponding δ-values on the y-axis. The first three lines show the bounds from the original Vuvuzela analysis, the last line our new bound for Gaussian noise (with better parameters).



(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 15: 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 conversation protocol). The figure depicts upper (red) and a lower bounds (blue) and Vuvuzela's privacy target (green, $\delta \leq 10^{-4}, e^\varepsilon \leq 2$).

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 ratio buckets 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 with the same variance, and we found evidence that the (ε, δ)-graph of a Laplace mechanism converges to the (ε, δ)-graph of a Gaussian mechanism with half the variance.

We encourage the application of our ratio buckets to other ADP mechanisms, such as to the optimal ADP mechanisms [9, 13] (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 between ADP of the Gaussian mechanism and ADP of the Laplace mechanism, as well as analyses probing the development of ADP provided by other noise distributions under composition interesting 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. Numpy.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. Springer, 2006.

[5] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our Data, Ourselves: Privacy Via Distributed Noise Generation. In *Advances in Cryptology - EUROCRYPT 2006*, pages 486–503. Springer, 2006.

[6] C. Dwork, M. Naor, T. Pitassi, and G. N. Rothblum. Differential Privacy Under Continual Observation. In *Proceedings of the 42th Annual ACM Symposium on Theory of Computing (STOC)*, pages 715–724. ACM, 2010.

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

[8] C. Dwork, K. Talwar, A. Thakurta, and L. Zhang. Analyze Gauss: Optimal Bounds for Privacy-preserving Principal Component Analysis. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC)*, pages 11–20. ACM, 2014.

[9] 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.

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

[11] T.-H. Hubert Chan, E. Shi, and D. Song. Private and Continual Release of Statistics. In *Automata, Languages and Programming. ICALP 2010*, pages 405–417. Springer, 2010.

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

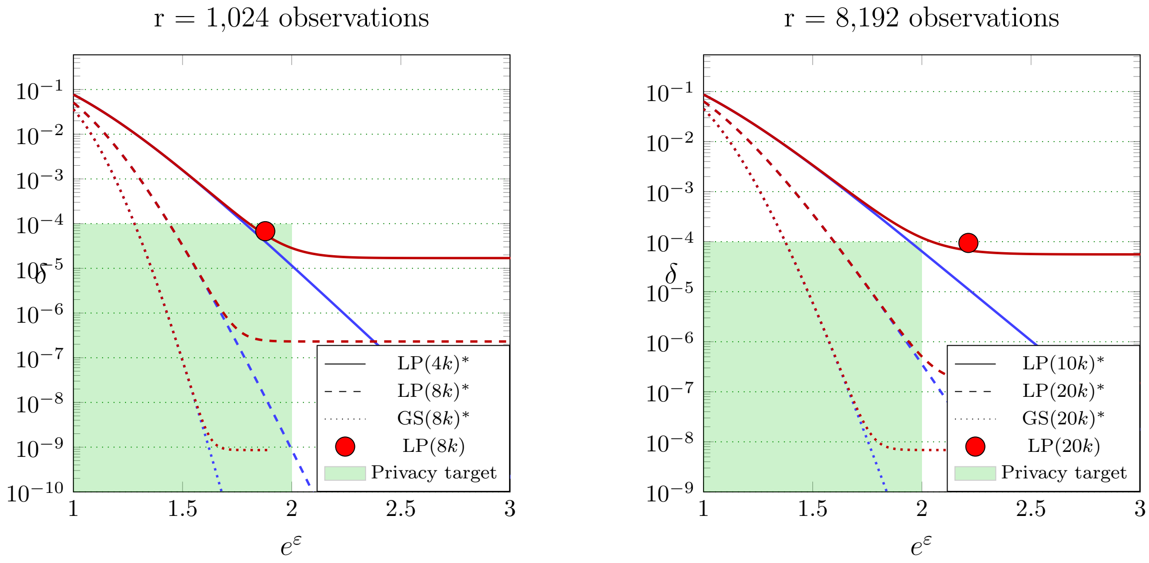
[13] K. Kalantari, L. Sankar, and A. D. Sarwate. Optimal differential privacy mechanisms under Hamming distortion for structured source classes. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 2069–2073. IEEE, 2016.

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

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

[16] 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.

[17] J. Tang, A. Korolova, X. Bai, X. Wang, and X. Wang. Privacy Loss in Apple’s Implementation of Differential Privacy on MacOS 10.12. *ArXiv e-prints*, 2017.



(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 16: 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, $\delta \leq 10^{-4}, e^{\varepsilon} \leq 2$).

[18] 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)*, pages 137–152. ACM, 2015.

A Appendix

A.1 Example calculation for [1]

In their paper, Abadi et al. cleverly combine methods to reduce the noise they require by a factor of q, which in one of their examples is $q = 0.01$. While the authors significantly improve the differential privacy bounds compared to prior work, their bound is not optimal. If we compare the (ε, δ) values for σ that their bound achieves with the results that we achieve for Gaussian noise, we get the following calculation.

We start by deriving values for c_1, c_2 from their example and we see that the closest possible values in their inequalities are $c_1 \approx 1.26$ and $c_2 \approx 1.485$. Using as an example our Gaussian noise with sensitivity one for 512 compositions, where we have $\sigma_{\text{us}} = \sqrt{2} \cdot 200$, we see that in order to apply their result, we need $\varepsilon < c_1 \cdot q^2 \cdot T \approx 1.26 \cdot 0.01^2 \cdot 500 \approx 0.063$. We use this value for ε and choose the appropriate δ output by our bucket distributions as $\delta \approx 0.01$. Consequently we get $\sigma \geq c_1 \cdot \frac{q\sqrt{T \cdot \log(1/\delta)}}{\varepsilon} = 1.485 \cdot \frac{0.01 \cdot \sqrt{500 \log(100)}}{0.0639} \approx 11.1515$. If we consider that the linear improvement of q to σ would also apply if we analyze the rest of the formula with our ratio buckets, we get $\sigma = q \cdot \sigma_{\text{us}} = 0.01 \cdot \sqrt{2} \cdot 200 \approx 2.8284$, which is significantly smaller (almost by a factor of 4) than their bound of $\sigma > 11.1515$.