One-Sided Bounded Noise: Theory, Optimization Algorithms and Applications

Hanshen Xiao
Purdue University / NVIDIA
West Lafayette, IN, USA
hsxiao@purdue.edu

Jun Wan
Five Rings
New York, NY, USA
junwanthu@gmail.com

Elaine Shi
Carnegie Mellon University
Pittsburgh, PA, USA
runting@cs.cmu.edu

Srinivas Devadas
Massachusetts Institute of Technology
Cambridge, MA, USA
devadas@mit.edu

Abstract

We investigate the optimal trade-off between utility and privacy using *one-sided* perturbation. Unlike conventional privacy-preserving statistical releases, randomization for obfuscating side-channel information is often constrained by infrastructure limitations. In practical scenarios, these constraints may only allow *positive* and *bounded* perturbations. For example, extending processing time or sending and storing dummy messages/data is typically feasible. However, implementing modifications in the opposite direction is challenging due to restrictions imposed by hardware capacity, communication protocols, and data management systems. In this paper, we establish the foundation of the positive noise mechanism within three semantic privacy frameworks: Differential Privacy (DP), Maximal Leakage (MaxL), and Probably Approximately Correct (PAC) Privacy. We then present a series of results that characterize or approximate the *optimal* one-sided noise distribution, subject to a second-moment budget and a bounded maximal magnitude. Building on this theoretical foundation, we develop efficient tools to solve the underlying optimization problems. Through experiments conducted in various scenarios, we demonstrate that existing techniques, such as Truncated Biased Laplace noise, are often suboptimal and result in excessive performance degradation. For instance, in an anonymous communication system with a $250K$ message budget, our optimized DP noise mechanism achieves a 21× reduction in dummy messages and an 18× reduction in dummy message latency overhead compared to traditional methods.[1]

CCS Concepts

• **Security and privacy** → Privacy-preserving protocols.

Keywords

Minimal Constrained Noise, Differential Privacy, Maximal Leakage, PAC privacy, Side-Channel Leakage

[1]All the proofs in this paper can be found in the full version https://hanshen-xiao.github.io/files/one_sided_noise.pdf.

CCS '25, Taipei, Taiwan

ACM ISBN 979-8-4007-1525-9/2025/10
https://doi.org/10.1145/3719027.3765110

ACM Reference Format:
Hanshen Xiao, Jun Wan, Elaine Shi, and Srinivas Devadas. 2025. One-Sided Bounded Noise: Theory, Optimization Algorithms and Applications. In . ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3719027.3765110

1 Introduction

Perturbation is one of the most general approaches to randomize a processing procedure for privatizing information disclosure. Various mathematical notions of privacy, including Differential Privacy (DP) [22, 23], Maximal Leakage (MaxL) [28] and PAC Privacy [46] have been established to formally quantify the attainable privacy guarantees through noise addition.

Most commonly-adopted noises are *symmetric*, which, to be more specific, are *unbounded* and *two-sided* centered around 0 (i.e., zero mean). Examples include standard Laplacian noise for pure ϵ-differential privacy (DP) [23], and Gaussian noise for approximate (ϵ, δ)-DP [22] and PAC Privacy [38, 46, 47].

However, many real-world applications impose specific constraints on the noise mechanisms. Due to application semantics, negative noise values may be impermissible. For instance, in anonymous communication systems, it is feasible to inject dummy messages, but removing existing messages could compromise functionality [11, 41]. Similarly, in encrypted database applications, one may pad the running time or the number of memory accesses needed to obscure the query and database contents [14], but reducing them could lead to errors or incorrect outputs. All of the above scenarios can be modeled using *one-sided bounded noise*, i.e., noise that is non-negative, has a positive mean (also referred to as bias), and lies within a bounded interval.

Take DP as an example. The most common practice for adding one-sided noise so far is to rely on a truncated and shifted Laplacian (or geometric) distribution [14, 41]. This approach involves shifting the mean of a standard Laplacian distribution to the positive (right) side until the negative (left) tail becomes sufficiently small, which will be truncated; we then redistribute the probability mass elsewhere after truncation. Unfortunately, this approach does *not* yield optimal error, either for a single disclosure setting or for multiple disclosures that require composition. One key result we will demonstrate is that, the optimal one-sided noise distribution is generally *asymmetric* and heavily *dependent* on the security parameters. This is dramatically different from the case of non-biased noise where *independent of* target DP security parameters, the optimal form

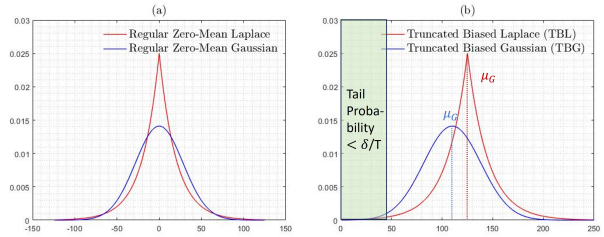


Figure 1: Illustration of zero-mean Laplace/Gaussian Mechanism and positive truncated biased Laplace/Gaussian Mechanism of the same noise variance. (1) Laplace, the taller (red) distribution, is less concentrated with a heavier tail compared to Gaussian, the wider (blue) distribution. (2) To produce the same (ϵ, δ)-DP guarantee, truncated Laplace noise requires a larger bias μ_G but less variance compared to truncated Gaussian noise. In general, the concentration of optimal DP one-sided noise needs to be carefully selected *depending on* the given (ϵ, δ) security parameters to balance the bias (the mean μ_G) and variance required, and the optimal noise form is generally *not* simply either a Laplace or Gaussian.

belongs to or can be closely approximated by some simple distribution class, such as staircase Laplace [27] or Gaussian [5]. Instead, *case-by-case* optimization is necessary for one-sided noise.

The starkly different landscapes are largely due to the different characteristics of the error produced by noise. Throughout this paper, we define the error of injected noise by the expectation of its square (second moment). For one-sided noise, both the bias (i.e., the positive mean) and variance of the noise distribution contribute to the error, unlike the case of unbiased noise where *only* its variance affects the error. The optimal distribution of one-sided noise thus needs to be carefully selected to balance both bias and variance simultaneously, and, unfortunately, simply shifting classic distributions, including staircase Laplace or Gaussian distributions, is generally *no* longer optimal as partially illustrated in Fig. 1, though they yield (asymptotic) optimal utility (error)-privacy trade-off in the regime of non-biased perturbation, .

In addition to the absence of theory to understand the fundamental gap in utility loss/error caused by the one-sided constraint, to the best of our knowledge, there is also *no* known framework to help privacy practitioners determine the optimal one-sided noise distribution. Note that for two-sided noise, especially in a *compositional* setting with multiple disclosures, many software toolboxes, including OpenDP [26] and Opacus [51], have been developed to automatically optimize noise parameters in standard DP implementations. A library for one-sided noise mechanisms, especially for different privacy metrics, is highly desirable.

1.1 Our Contributions

In this paper, we not only lay the foundation for characterizing the minimal one-sided bounded noise under three semantic privacy definitions—Differential Privacy (DP), Maximal Leakage (MaxL), and PAC Privacy—but also introduce an open-source library, `lib-1sided-noise` [2] to assist privacy practitioners in their tasks. In general, the optimal one-sided noise does *not* have a closed-form expression, and the solutions vary across different privacy definitions. To address this, we develop optimization algorithms separately with efficient implementations in our library. These enable automatic determination or approximation of the optimal noise under the selected privacy metric and budget. Additionally, we provide comprehensive comparisons of the semantic interpretations of privacy definitions to guide practitioners in selecting the most appropriate metrics for addressing diverse privacy concerns across applications (see Section 2.4). We summarize our theoretical and algorithmic contributions below:

Differential Privacy For a single release (without composition) with (ϵ, δ)-DP guarantee, we present the optimal positive, bounded noise distribution (Theorem 1). For multiple releases, we present a new and tight composition accounting by generalizing the Rényi divergence (Theorem 2) and transforming the determination of minimal positive noise under T-fold composition into a constrained optimization. Finally, we show how to iteratively apply convex optimizations (Algorithm 1) to optimize the noise distribution (Theorem 3) to approximate the optimal solution.

Maximal Leakage We prove that the optimal positive perturbation strategy with minimal cost/overhead to produce $\log(v)$-MaxL, for an arbitrary real number $v \geq 1$, is a linear interpretation of the optimal deterministic perturbation schemes to produce $\log(\lfloor v \rfloor)$ and $\log(\lceil v \rceil)$ MaxL, respectively (Theorem 4). We then apply dynamic programming to build the *first* efficient algorithm (Algorithm 2) to find the provably optimal perturbation in polynomial time.

PAC Privacy We study the PAC Privacy bound for black-box processing given the output variance. We prove given the second moment budget and a maximal magnitude restriction, the optimal positive noise must be within a family of truncated Gaussian (Theorem 5) and the problem is reduced to optimize the mean and variance of the Gaussian being truncated. We present an efficient optimization algorithm that executes a series of simple iterative binary searches (Algorithm 3).

1.2 Concrete Results on Applications

1.2.1 Network Traffic Leakage. Anonymous communication while hiding who is communicating with whom has received significant attention. It is known that even if messages are encrypted, metadata that reveals active users can be recovered through network traffic analysis [31, 41]. A common privatization mechanism is to ask users to send dummy messages [30], which is essentially a positive perturbation to the packet volume observed by the adversary. Its privacy guarantee has been measured from a DP point of view in [39, 41]. On the other hand, if one adopts dropping messages or partitioning messages into multiple communications to produce a negative obfuscation, it can break soundness unless the system has redundancy built-in, which itself can be expensive.

Compared to truncated Laplace noise adopted in [8, 14, 34, 41], we show the scale of optimized one-sided DP noise can be *orders* of magnitude smaller in practice: for *Vuvuzela* anonymous communication [41], we achieve 21× less dummy messages with 18× less dummy message latency overhead for a 250K message budget.

[2] https://github.com/Hanshen-Xiao/lib-1sided-noise

1.2.2 Mitigating Cache-Timing Attacks. The execution of cryptographic algorithms or operations on different inputs takes different times and, in practice, it is challenging to write high-performance constant-time software for general-purpose computers [7]. This additional data-dependent timing information could leak cryptographic secrets; in some applications, timing information can be simply characterized by hits and misses to a cache [29]. Bernstein in [7] presented concrete cache-timing attacks to fully recover an Advanced Encryption Standard (AES) secret key. More recently, for OpenSSL's constant-time code, a cache-timing attack on RSA key generation was developed by exploiting a vulnerable code path [2]. Even in Trusted Execution Environments (TEEs), through cache-timing and speculation attacks, Bulck et al. proposed *Foreshadow* to break Intel Software Guard eXtensions (SGX) [40]. In practical execution, it is easy to turn cache hits into misses, but not vice versa. To obfuscate timing and cache information and provably mitigate attacks, the corresponding modification must be positive.

In this paper, we present the first formal analysis with optimal one-sided noise to obfuscate the cache-timing leakage in a 256-bit AES secret key generation from S-boxes [4] under both MaxL and PAC Privacy. In particular, we show a small random positive noise with 1 dummy miss in expectation can provably ensure a negligible adversarial success rate ($< 2^{-245}$) to correctly recover a 256-bit secret key (Fig. 4).

2 Preliminaries

In this section, we formally introduce three semantic and rigorous privacy definitions: Differential Privacy (DP) [21, 23], Maximal Leakage (MaxL) [28], and PAC Privacy [46]. At a high level, the problem of information leakage control can be described by the following generic model: for some sensitive data/input $X \in \mathcal{X}^*$ and some processing function $\mathcal{F} : \mathcal{X}^* \to \mathcal{Y}^*$, the output $\mathcal{F}(X)$ represents the release/leakage. The goal of privacy preservation is to randomize or modify the original processing function $\mathcal{F}$ into a version $\mathcal{M}$ such that provided the randomized or noisy output $\mathcal{M}(X)$, the adversary cannot implement meaningful inference on the sensitive input X. The following privacy definitions offer formal languages to quantify such hardness from different angles. In Section 2.4, we include a comparison regarding their applicability and the underlying operational challenges in practice.

2.1 Differential Privacy

Definition 1 (Differential Privacy [22]). *Given a dataset universe $\mathcal{X}^*$, we say that two datasets $X, X' \subseteq \mathcal{X}^*$ are adjacent, denoted as $X \sim X'$, if X can be obtained by replacing one datapoint in X', i.e., $X = (X'/x') \cup x$. A randomized algorithm $\mathcal{M}$ is said to be (ϵ, δ)-differentially-private (DP) if for any pair of adjacent datasets X, X' and any event set Y in the output domain of $\mathcal{M}$*

$$\mathbb{P}(\mathcal{M}(X) \in Y) \leq e^{\epsilon} \cdot \mathbb{P}(\mathcal{M}(X') \in Y) + \delta. \tag{1}$$

(ϵ, δ)-DP enjoys an intuitive interpretation where e^{ϵ} and δ represent a multiplicative and an additive term, respectively, to capture the worst-case divergence/difference between the likelihood functions produced by two arbitrary adjacent datasets X and X'. From a hypothesis testing perspective, small ϵ and δ will imply either a large Type I or Type II error [20]. In practice, Gaussian and Laplace mechanisms are the workhorses to randomize a processing function for $\epsilon(,\delta)$-DP guarantee [24]: the scale of noise is calibrated to the *sensitivity*, i.e., the maximal possible change to the output when one arbitrarily replaces a single datapoint. Unfortunately, tight sensitivity is in general NP-hard to compute [50]. Thus, an important concept in DP research is *composition*, which captures the cumulative privacy risk from multiple releases and plays a key role to privatize algorithms in practice. A complicated algorithm with intractable sensitivity, such as DP-SGD, is usually decomposed into multiple, relatively simpler suboperations with tractable/bounded sensitivity, such as gradient mean estimation of a batch of samples [1, 43, 48, 49]; one can then perturb the intermediate outcomes from each suboperation, assuming that they are released, and derive an upper bound by composing the privacy loss of the release from each iteration, as formalized in the following proposition.

Proposition 1 (Advanced Composition [25]). *For any $\epsilon_0 > 0$ and $\delta_0 \in (0, 1)$, the class of (ϵ_0, δ_0)-differentially private mechanisms satisfies $(\epsilon, T\delta_0 + \tilde{\delta})$-differential privacy under T-fold adaptive composition, where, for any $\tilde{\delta} > 0$,*

$$\epsilon = \sqrt{2T \log(1/\tilde{\delta})} \cdot \epsilon_0 + T\epsilon_0(e^{\epsilon_0} - 1). \tag{2}$$

When $\epsilon_0 = o(1/\sqrt{T})$, the latter term $T\epsilon_0(e^{\epsilon_0} - 1)$ is $o(1)$ and the advanced composition (2) roughly states that ϵ scales in $O(\sqrt{T})$ under T-fold composition, which is general and only counts on the (ϵ_0, δ_0)-DP guarantee per release.

2.2 Maximal Leakage

Similar to DP which considers an input-independent guarantee, Maximal Leakage (MaxL) is another operationally-interpretable definition, which measures how much more likely the adversary can identify the true input given the release. Let $U \in \mathcal{U}^*$ denote the user's secret, $X \in \mathcal{X}^*$ denote some intermediate parameter of a processing mechanism $\mathcal{M} : \mathcal{X}^* \to \mathcal{Y}^*$, whose input and output domains are $\mathcal{X}^*$ and $\mathcal{Y}^*$, respectively, and $Y \in \mathcal{Y}^*$ denote the output of $\mathcal{M}$. Here, $\mathcal{U}^*$, $\mathcal{X}^*$ and $\mathcal{Y}^*$ are all assumed to be finite sets, which can be generalized to the continuous case [28]. Clearly, U, X and Y form a Markov Chain, denoted as $U - X - Y$; the formal definition of MaxL is given below.

Definition 2 (Maximal Leakage [9, 28]). *Let $Adv : \mathcal{Y}^* \to \mathcal{U}^*$ be an arbitrary algorithm that the adversary applies to recover the secret U from the observation on Y. MaxL with respect to the processing procedure $U - X - Y$ is defined as*

$$\mathcal{L}(X \to Y) = \sup_{U:U-X-Y} \log \frac{\sup_{Adv} \Pr(U = Adv(Y))}{\max_u \Pr(U = u)}, \tag{3}$$

which is known to enjoy an equivalent form [3, 28, 33]

$$\mathcal{L}(X \to Y) = \log \sum_{y \in \mathcal{Y}^*} \max_{x \in \mathcal{X}^*} \mathbb{P}(Y = y | X = x). \tag{4}$$

Compared to DP with a particular focus on *individuals*, the privacy concern of MaxL is regarding the full reconstruction over the *entire* input U, where (3) upper bounds the *multiplicative* gain to adversary's posterior knowledge on the secret U after observing the release Y. To be more specific, if a mechanism $\mathcal{M}$ satisfies $\log(v)$-MaxL, then the ratio between the optimal *posterior* chance that an adversary can correctly identify the true input,

$\sup_{Adv} \Pr(U = Adv(Y))$, and the optimal *a priori* success rate that the adversary can identify the true input, $\sup_{\tilde{U}} \Pr(U = \tilde{U})$, is bounded by e^{υ} for *any* possible prior distribution of U.

Additionally, the non-adaptive composition of MaxL enjoys a simple summation form [28]. In the same setup $U - X - \bar{Y} = (Y_1, Y_2)$, let Y_1 and Y_2 be the releases from two mechanisms which are independent conditional on X, then we have

$$\mathcal{L}(X \to \bar{Y}) \le \mathcal{L}(X \to Y_1) + \mathcal{L}(X \to Y_2). \tag{5}$$

2.3 PAC Privacy

From an input-independent perspective, Differential Privacy (DP) and Maximal Leakage (MaxL) measures privacy risk/loss in terms of the worst-case distinguishability and the multiplicative gain of successfully identifying the secret, respectively. However, both DP and MaxL require *white-box* algorithmic analysis—such as sensitivity or output likelihood—of the processing function $\mathcal{F}$ to derive provable privacy solutions. This reliance makes it challenging to handle (side-channel) leakage from complicated circuits/processors or involved protocols which generally lacks closed form; meanwhile, membership and identification may also not tightly capture broader inference problems, such as (partial) reconstruction attacks.

In contrast, PAC Privacy supports *black-box* privatization. A noise solution can be automatically determined to resist *any* target adversarial inference, and PAC Privacy analysis only requires black-box access to the underlying secret generation and processing function $\mathcal{F}$ [46]. Conceptually, PAC Privacy challenges an adversary to return a *satisfactory* estimation $\tilde{X}$ of the sensitive input X and quantifies the posterior success probability. The criterion of satisfactory reconstruction reflects the level of leakage deemed unacceptable by the secret holder, where, for example, $\tilde{X}$ approximates the salary attribute of a record X with error smaller than 1000, or predicts at least 200 bits correctly of a 256-bit secret key X. The formal definition is given below.

Definition 3 ($(\delta_\rho, \rho, \mathsf{D})$ PAC Privacy [38, 45–47]). *For a processing mechanism $\mathcal{M} : \mathcal{X}^* \to \mathcal{Y}^*$, data distribution* D*, and an inference criterion function $\rho(\cdot,\cdot)$, we say $\mathcal{M}$ satisfies $(\delta_\rho, \rho, \mathsf{D})$-PAC Privacy if the following experiment is impossible:*

A user generates data X from distribution D *and sends $\mathcal{M}(X)$ to an adversary. The adversary who knows* D *and $\mathcal{M}$ is asked to return an estimation $\tilde{X}$ on X such that with probability at least $(1-\delta_\rho)$, $\rho(\tilde{X}, X) = 1$.*

In Definition 3, the probability is based on the randomness in both secret generation $X \leftarrow \mathsf{D}$ and the mechanism $\mathcal{M}$. The criterion of satisfactory estimation is captured by the indicator function ρ where $\rho(\tilde{X}, X) = 1$ if and only if $\tilde{X}$ successfully recovers the targeted feature of X. Continuing with the previous examples, we can define $\rho(\tilde{X}, X) = 1$ iff the estimation error of $\tilde{X}$ on the salary attribute smaller than 1000 or if X and $\tilde{X}$ collide in least 200 bits.

Operationally, to upper bound $(1-\delta_\rho)$, [46] studies the difference between the *optimal* prior and posterior success rate in f-divergence [37]. Let $(1-\delta_{o,\rho})$ denote the optimal *a priori* success rate, i.e., the best chance that an adversary can return a satisfied estimation $\tilde{X}$ such that $\rho(\tilde{X}, X) = 1$ *before* observing the release $\mathcal{M}(X)$:

$$1 - \delta_{o,\rho} = \arg_{X' \in \mathcal{X}^*} \Pr_{X \leftarrow \mathsf{D}}(\rho(X', X) = 1). \tag{6}$$

[46] introduces two Bernoulli variables $\mathbf{1}_{\delta_\rho}$ and $\mathbf{1}_{\delta_{o,\rho}}$, where $\Pr(\mathbf{1}_{\delta_\rho} = 1) = (1-\delta_\rho)$ and $\Pr(\mathbf{1}_{\delta_{o,\rho}} = 1) = (1-\delta_{o,\rho})$, respectively, and considers the f-divergence between these two Bernoulli distributions,

$$\Delta_f^\rho = \mathcal{D}_f(\mathbf{1}_{\delta_\rho} \| \mathbf{1}_{\delta_{o,\rho}}) = \delta_{o,\rho} f(\frac{\delta_\rho}{\delta_{o,\rho}}) + (1-\delta_{o,\rho}) f(\frac{1-\delta_\rho}{1-\delta_{o,\rho}}), \tag{7}$$

where $f(\cdot)$ can be an arbitrary convex function. It is noted that given the data distribution D and the inference task ρ of interest, the optimal prior rate $(1-\delta_{o,\rho})$ is determined. Thus, an upper bound of Δ_f^ρ in (7) combined with a lower bound of the prior rate $(1-\delta_{o,\rho})$ will lead to an upper bound of the target posterior success rate $(1-\delta_\rho)$. In particular, as a special case of Theorem 1 in [46], when we select the f-divergence to be the KL-divergence, i.e., by selecting $f(t) = t \log t$ in (7), Δ_{KL}^ρ is shown to be bounded by the well-known mutual information [17], as described below.

Proposition 2 ([46]). *For arbitrary ρ and input distribution* D*,*

$$\Delta_{KL}^\rho = \delta_\rho \log(\frac{\delta_\rho}{\delta_{o,\rho}}) + (1-\delta_\rho) \log(\frac{1-\delta_\rho}{1-\delta_{o,\rho}}) \le \mathsf{MI}(X; \mathcal{M}(X)), \tag{8}$$

where $\mathsf{MI}(X; \mathcal{M}(X))$ is the mutual information between X and $\mathcal{M}(X)$.

Stemmed from (8), [46] shows how to automatically determine the minimal perturbation e for a black-box processing function $\mathcal{F}$ based on high-confidence estimation on the (co-)variance of $\mathcal{F}(X)$, which ensures that its noisy version $\mathcal{M}(X) = \mathcal{F}(X) + \mathsf{e}$ satisfies provable PAC Privacy guarantees.

2.4 Proper Selection of Privacy Metric

In the previous subsections, we formally introduced three privacy definitions. In practice, before selecting a privacy metric, one should first determine the privacy concern of interest – which part of input data X we aim to protect. The objective could vary from an attribute, to a data point, to relationship between datapoints. DP puts a particular focus on individual privacy and can provide meaningful guarantees especially when the release is an aggregation of multiple individuals, for example, the network traffic when a set of users communicate [41] or the memory pattern when one accesses a database [14], and the goal here is to prevent the adversary from inferring who is talking or which file is accessed. As a comparison, MaxL is not restricted to only make an individual indistinguishable, but also to bound the posterior advantage for an adversary to correctly identify the true input. Thus, MaxL can capture the privacy leakage (adversarial reconstruction hardness) with respect to the *entire* input and can be applied to study the leakage from, for example, processing time or power consumption of a specific program [19], where individual or attribute privacy is *not* meaningful or well-defined. Compared to DP and MaxL, PAC Privacy offers the most general framework to probabilistically describe inference hardness in recovering any specified related information regarding the secret X. However, PAC Privacy requires an *a priori* setup (secret entropy), which is different from the input-independent guarantee in DP and MaxL. In the following, we demonstrate the applicability and operational challenges of the three frameworks.

To randomize a processing function $\mathcal{F}$ to satisfy DP guarantees, one needs to first bound sensitivity, the worst-case change $(\sup_{X \sim X'} \|\mathcal{F}(X) - \mathcal{F}(X')\|$ for arbitrary two adjacent datasets $X \sim X')$ of the output when one arbitrarily replaces a datapoint. As

mentioned before, tight sensitivity is intractable in many practical applications, and, usually, one needs to introduce some artificial control, such as clipping and decomposition [1], to produce a sensitivity bound. As for MaxL, sensitivity analysis is not necessary, but MaxL still requires the knowledge of likelihoods across all input selections. Different from DP and MaxL, which cannot view the underlying processing $\mathcal{F}$ as a black box, PAC Privacy enables automated privatization for general inference hardness, not only restricted to distinguishability or identification. However, PAC Privacy requires that the secret distribution D is given or one can repeatedly sample from D. Thus, PAC Privacy is more suitable for statistical data processing or secret key protection, where the input has a clear form of entropy. For example, in protecting an l-bit secret key from cache timing attacks [7, 19, 40], the distribution D of a random secret key X is a uniform distribution over $\{0,1\}^l$ and thus PAC Privacy is easily applicable. However, when our secret X cannot be sampled or does not enjoy tractable entropy, for example, the messages in anonymous communication, input-independent guarantees becomes the only known feasible solution.

3 A Lesson from Biased Noise —Tradeoff between Mean, Variance and Concentration

In this section, we provide some intuition on the following two important questions: a) why existing positive noise constructions could be sub-optimal, and b) how to construct the optimal positive perturbation. We mainly focus on DP positive noise in this section, but the implications of the results are general, which instruct our following study on the optimal one-sided noise.

In prior works on mitigating side-channel leakage with DP guarantees, Truncated Biased Laplace (TBL) noise [6, 8, 15, 35] and its discrete version, Truncated Geometric noise [13, 52] are among the most-commonly used perturbations. We first take the continuous TBL noise as the example; see definition below.

Definition 4 (Truncated Biased Laplace Noise [6]). *Given parameters $\mu_L > 0$, $\lambda_L > 0$ and $R > 0$, a (μ_L, λ_L, R) Truncated Biased Laplace (TBL) truncates a Laplacian distribution*

$$\mathsf{Lap}_{\mu_L,\lambda_L}(z) = \frac{1}{2\lambda_L} exp(-\frac{|z-\mu_L|}{\lambda_L}) \tag{9}$$

on range $[0, R]$. The resulted noise e has a probability distribution

$$\mathsf{P}(\mathsf{e}=z) = \frac{1}{Z_{\mu_L,\lambda_L,R}} exp(-\frac{|z-\mu_L|}{\lambda_L}) \cdot \mathbf{1}_{0\le z\le R}. \tag{10}$$

$\mathbf{1}_{0\le z\le R}$ is an indicator which equals 1 when $z \in [0, R]$, otherwise 0. $Z_{\mu_L,\lambda_L,R} = \int_0^R exp(-\frac{|z-\mu_L|}{\lambda_L})dz$ is the normalization parameter.

The following lemma describes the (μ, λ) selection of TBL noise such that it can produce an (ϵ, δ)-DP guarantee.

Lemma 1 (Parameter of Positive Laplace Noise)**.** *Suppose a processing function $\mathcal{F} : \mathcal{X}^* \to \mathbb{R}$ such that for an arbitrary adjacent dataset pair $X \sim X'$, $|\mathcal{F}(X) - \mathcal{F}(X')| \le \mathsf{s}$, i.e., the sensitivity of $\mathcal{F}$ is bounded by s. Then, if we select $\lambda_L = \mathsf{s}/\epsilon$, $\mu_L \ge \mathsf{s} + \frac{\mathsf{s}}{\epsilon} \cdot \log \frac{1}{2\delta(1-e^{-\mu_L\cdot\epsilon/\mathsf{s}})}$, and $R = 2\mu_L$, such a (μ_L, λ_L, R)-TBL perturbation ensures (ϵ, δ)-DP.*

Intuitively, TBL noise can be viewed as that we perform the following modifications to a standard zero-mean Laplace distribution $\mathsf{Lap}_{0,\lambda}$, defined in (9). First, we shift the $\mathsf{Lap}_{0,\lambda}$ uniformly by μ into $\mathsf{Lap}_{\mu,\lambda}$; second, we truncate its support domain from (∞, ∞) to $[0, R]$ and normalize the remaining over $[0, R]$. Started from TBL noise, we have several remarks on positive noise mechanisms and the resultant DP guarantees, compared to the regular zero-mean noise mechanism:

(1) **ϵ-DP is impossible and a failure probability is necessary.** For an arbitrary positive noise e and an arbitrary deterministic processing function, the respective support sets[3] of the distributions of $\mathcal{F}(X) + \mathsf{e}$ and $\mathcal{F}(X') + \mathsf{e}$, for $X \sim X'$, $\mathcal{F}(X) \ne \mathcal{F}(X')$, *cannot be identical.* There always exists some subset O such that $\Pr(\mathcal{F}(X) + \mathsf{e} \in O) > 0$ while $\Pr(\mathcal{F}(X') + \mathsf{e} \in O) = 0$. Once the outputs fall within O, the adversary can perfectly distinguish the input between X and X', and thus an additive failure rate in a positive noise mechanism is necessary. In TBL, to ensure such a failure probability bounded by δ, we need to select a large enough shift/bias μ_L such that the tail probability between $[0, \mathsf{s}]$ or $[R - \mathsf{s}, R]$ is bounded by δ, where s is the sensitivity bound.

(2) **Heavier Utility Loss:** Consider the utility loss captured by the second moment of the injected noise e,

$$\mathbb{E}[\mathsf{e}^2] = \underbrace{\big(\mathbb{E}[\mathsf{e}]\big)^2}_{\text{mean}} + \underbrace{\mathbb{E}\big[\big(\mathsf{e}-\mathbb{E}[\mathsf{e}]\big)^2\big]}_{\text{variance}}. \tag{11}$$

Given the positive requirement of e, $\mathbb{E}[\mathsf{e}]$, the mean of e, must be non-zero, and thus compared to the regular zero-mean noise mechanism, in general, we need to pay an additional utility loss determined by the square of the mean, i.e., $\big(\mathbb{E}[\mathsf{e}]\big)^2$, proportional to $(\mu_L)^2$ in the case of TBL noise.

From Lemma 1, we know that TBL noise behaves as a sufficient method to produce (ϵ, δ)-DP. A natural question is whether TBL is optimal. Before we give a complete answer in the next section, we first show a more intuitive negative answer that TBL is *not* optimal in producing a tight utility-privacy tradeoff under composition. When we consider T-fold composition of a noisy mechanisms using TBL, the cumulative failure probability from the tail scales with T by union bound. Thus, to ensure (ϵ, δ) after T compositions, by advanced composition (Proposition 1) with $\tilde{\delta} = \delta/2$, it suffices to ensure (ϵ_0, δ_0)-DP in each round, where

$$\epsilon_0 = O(\frac{\epsilon}{\sqrt{2T\log(2/\delta)}}), \text{ and } \quad \delta_0 = \frac{\delta}{2T}. \tag{12}$$

Thus, to ensure (ϵ_0, δ_0)-DP by TBL noise, from Lemma 1, we may select $\lambda_L = \frac{\mathsf{s}}{\epsilon_0}$ and $\mu_L = O(\lambda_L \cdot \log \frac{T}{2\delta} + \mathsf{s})$. Consequently, the second moment of the constructed TBL noise is

$$\mathbb{E}[e^2] = O\big((\mu_L)^2 + (\lambda_L)^2\big) = O\big(\frac{T\log(1/\delta)}{\epsilon^2} \cdot (\underline{\log^2(T/\delta)} + 1)\big), \tag{13}$$

for constant s. It is worthwhile noting that the mean/bias of the TBL noise e is $O(\log T/\delta)$ times larger than its standard deviation, as underlined in (13). This matches our intuition that given an $O(e^{-z})$ decaying rate of Laplace distribution, only after an $O(\lambda_L \log 1/(\delta/T))$ length distance from its mean, can we ensure its noise tail is small enough in $O(\delta/T)$. As a comparison, we may similarly consider the Truncated Biased Gaussian (TBG) noise, as an analog of TBL.

[3] Throughout the paper, we use support set to represent the domain for a distribution, over which there is non-zero probability density/mass.

Definition 5 (Truncated Biased Gaussian). *Given $\mu_G > 0$ and $\lambda_G > 0$, the probability density function of a (μ_G, λ_G, R) Truncated Biased Gaussian (TBG) noise* e *is defined as*

$$\mathsf{P}(\mathsf{e} = z) = \mathbf{1}_{0 \le z \le R} \cdot \frac{1}{Z_{\mu_G,\lambda_G,R}} exp(-\frac{(z-\mu_G)^2}{2\lambda_G^2}), \tag{14}$$

where $Z_{\mu_G,\lambda_G,R} = \int_0^R exp(-\frac{(z-\mu_G)^2}{2\lambda_G^2})dz$ is for normalization.

If we adopt TBG noise to produce (ϵ, δ)-DP under T compositions, similarly it suffices to select $\lambda_G = O(\frac{\sqrt{T\log(1/\delta)}}{\epsilon})$, $\mu_G = O(\sqrt{\log(T/\delta)}\lambda_G)$ and $R = 2\mu_G$. Then,

$$\mathbb{E}[e^2] = O\big((\mu_0)^2 + (\lambda_0)^2\big) = O\big(\frac{T\log(1/\delta)}{\epsilon^2} \cdot (\underline{\log(T/\delta)} + 1)\big). \tag{15}$$

Comparing (13) with (15), we observe the following:

Square of Mean Larger than Variance: In both positive noise mechanisms with either TBL or TBG, from (13) and (15), to ensure a small tail in a scale δ/T, the bias parameter $\mu_L(\mu_G)$ needs to be a polynomial of $(\log(T/\delta))$ times larger than their standard deviation $\sqrt{T\log(1/\delta)}/\epsilon$, required by the regular (zero-mean) Laplace / Gaussian mechanism to produce the same (ϵ, δ) under T-fold composition. Such a gap *cannot* be simply mitigated by a noise of larger variance: the bias $\mu_L(\mu_G)$ scales with the standard deviation, controlled by $\lambda_G(\mu_G)$, and a larger variance only makes the distributional decay slower, requiring an even larger mean/bias.

Concentration vs. Variance: A closer look at the underlined terms in both (13) and (15)—which capture how many times the mean (or bias) exceeds the standard deviation—reveals that the bias of the TBG noise is smaller than that of the TBL noise by a factor of $O(\log(T/\delta))$. This arises from the fact that the Gaussian distribution is more concentrated. Specifically, the tail of the Gaussian decays at a rate of $O(e^{-z^2})$, compared to $O(e^{-z})$ for the Laplace distribution. This faster decay enables the Gaussian mechanism to use a smaller bias while still achieving the desired tail bound. We illustrate this in Fig. 1 (on page 2).

Indeed, if we expand all constants in (13) and (15), we find that the TBG mechanism requires a larger noise variance than the TBL mechanism. This aligns with our intuition: to achieve the same statistical divergence with a faster-decaying noise distribution, a larger variance is necessary to compensate.

The above two observations suggest that the optimal positive noise distribution requires a careful tradeoff between mean (bias), variance, and concentration. To close the utility-loss gap relative to zero-mean noise, we seek a distribution with sufficient concentration to keep the bias small, while still providing a sharp tail bound to maintain a low failure probability (at most δ/T), guarding against distinguishability via differences in support. At the same time, the distribution must not be overly concentrated, which would result in an excessively large variance.

4 Positive Noise for Differential Privacy

In this section, we systematically study the optimal positive noise mechanism for Differential Privacy (DP). We consider a discrete processing function $\mathcal{F} : \mathcal{X}^* \to \mathcal{Y}^* \in \mathbb{Z}$, whose sensitivity is 1, i.e., for two arbitrary adjacent datasets $X \sim X' \in \mathcal{X}^*$, $\sup_{X\sim X'} |\mathcal{F}(X) - \mathcal{F}(X')| \le 1$. Our goal is to determine the optimal positive noise distribution of e over a bounded set $[0 : R] = \{0, 1, \cdots, R\}$ such that its second moment $\mathbb{E}[\mathsf{e}^2]$ is minimal while $\mathcal{F}(X) + e$ satisfies an (ϵ, δ)-DP guarantee (under T-fold composition).

Remark 1 (General Sensitivity)**.** *In the above setup we normalize the sensitivity to be 1, while all our following results on DP can be generalized to arbitrary integer sensitivity s by leveraging the idea of* group DP*: if our noise mechanism satisfies (ϵ, δ)-DP with sensitivity being 1, then for the case of sensitivity s, its corresponding DP guarantee scales to $(\epsilon/s, O(\delta/s))$ [24].*

4.1 Optimal Positive Noise for a Single Release

We begin by examining the optimal positive noise mechanism for (ϵ, δ)-DP in a single release scenario *without* composition ($T = 1$). We will prove the following more general conclusion: given an (ϵ, δ)-DP budget, the optimal noise distribution that minimizes the k-th moment for *any* $k \in \mathbb{Z}^+$ is *identical* and has a *closed-form* expression. Furthermore, as $R \to \infty$, meaning that even if there is no bounded restriction for the noise e, the optimal positive noise distribution remains inherently bounded. To be formal, we use $\mathsf{P}_e = \{p_0, p_1, \cdots\}$ to denote the noise distribution where $p_i = \Pr(e = i)$. The k-th moment of e is defined as $\mathbb{E}[e^k] = \sum_{i=0}^{\infty} i^k \cdot p_i$.

Theorem 1 (Optimum for Single Release)**.** *Given a processing function $\mathcal{F}$ of sensitivity 1, among all possible distributions of a positive noise* e *over $[0, +\infty)$ which ensure an (ϵ, δ)-DP guarantee of the noisy version $\mathcal{F}(\cdot) + \mathsf{e}$, the following distribution with probability mass function given in (1) below,*

$$p_i = \begin{cases} \delta \cdot e^{\epsilon i} & \text{if } i < \omega \\ \delta \cdot c \cdot e^{\epsilon(2\omega - i)} & \text{if } \omega \le i \le \omega', \end{cases} \tag{16}$$

is optimal in a sense that it achieves the minimal k-th moment, for any *positive integer k. Here, ω' is either $2\omega - 1$ or 2ω, and $c \in [e^{-2\epsilon}, 1]$ is for normalization such that the sum of p_i equals 1. Here, ω is a turning point, defined as*

$$\omega = \frac{1}{\epsilon} \cdot \log(\frac{2}{e^\epsilon + 1} + \frac{e^\epsilon - 1}{\delta(e^\epsilon + 1)}). \tag{17}$$

The proof of Theorem 1 is in Appendix B. Theorem 1 shows that the optimal distribution form is *identical*, in a form of (16), regardless of k. We provide some insights on how the parameters ω, ω' and c are determined. For given ω, we consider the following sequence $\{p_i(\omega)\}$ in a staircase:

$$p_i(\omega) = \begin{cases} \delta \cdot e^{\epsilon i} & \text{if } i < \omega \\ \delta \cdot e^{\epsilon(2\omega - i)} & \text{if } \omega \le i \le 2\omega. \end{cases} \tag{18}$$

(18) is the ideal noise shape to ensure an ϵ-multiplicative difference given sensitivity 1. To further ensure $p_i(\omega)$ is a valid probability mass, we need to select ω such that $S(\omega) = \sum_{i=0}^{2\omega} p_i(\omega) = 1$. However, $S(\omega) = 1$ may not have integer solutions, and we address this by setting $\omega = \min\{\omega' \in \mathbb{Z}^+ \mid S(\omega') \ge 1\}$, and introduce a scaling parameter $c \in (0, 1]$ to normalize p_i for $i \in [\omega, 2\omega]$. If $c < e^{-2\epsilon}$, the (ϵ, δ)-DP guarantee can no longer be maintained. In such cases, we adjust the upper bound of i from 2ω to $2\omega - 1$, ensuring that $c \ge e^{-2\epsilon}$.

As a summary, Theorem 1 suggests that given a range restriction R, either no noise distribution within $[0, R]$ can ensure the (ϵ, δ)-DP requirement, or the optimal P_e *must* be supported on $[0, R_0]$ for some $R_0 \le 2\omega \le R$, with ω defined in (17).

In Table 1 and 2, we include numerical results on the second moment of the TBL noise and the optimal noise (from Theorem 1) to achieve the same DP guarantees in various setups. Generally speaking, for stronger DP guarantees (smaller ϵ and δ), Theorem 1 brings more significant improvement.

Table 1: Comparison between the expected square (second moment) of Truncated Bounded Laplace (TBL) Noise and the optimal noise (Theorem 1) for various $(\epsilon, \delta = 10^{-4})$ DP guarantees for a single release.

	$\epsilon = 0.5$	$\epsilon = 1$	$\epsilon = 2$	$\epsilon = 4$	$\epsilon = 8$
TBL	265.5	83.0	27.1	9.8	4.3
Thm 1	244.5(-8%)	68.9(-17%)	17.0(-37%)	4.1(-58%)	1.0(-77%)

Table 2: Comparison between the expected square (second moment) of Truncated Bounded Laplace (TBL) Noise and the optimal noise (Theorem 1) for various $(\epsilon, \delta = 10^{-6})$ DP guaranteesfor a single release.

	$\epsilon = 0.5$	$\epsilon = 1$	$\epsilon = 2$	$\epsilon = 4$	$\epsilon = 8$
TBL	648.1	187.7	56.3	18.3	6.9
Thm 1	608.1(-6%)	171.5(-8%)	37.7(-33%)	9.1(-50%)	1.1(-84%)

4.2 Hybrid Rényi DP (HRDP)

In the previous section, we have studied the optimality of positive noise for (ϵ, δ)-DP in a *single* iteration. The analysis becomes more involved when we need to further consider composition. Although we introduced advanced composition of (ϵ, δ)-DP in Proposition 1 but we need to mention that (2) is not perfectly tight (both in constants and asymptotically, if we do not ignore the logarithm term [20]). This is especially true when we have additional information on the output distribution of $\mathcal{M}(X)$. For example, when the noise is Gaussian [1], it is known that the (ϵ, δ) metric does not fully characterize the statistical difference between $\mathcal{M}(X)$ and $\mathcal{M}(X')$ to produce the tightest composition. To this end, more involved DP definitions or accounting methods are proposed, such as zero-concentrated DP [10] and Rényi DP [32], which are both established based on Rényi divergence. We formally define Rényi DP as follows.

Definition 6 (Rényi Differential Privacy [32]). *A randomized algorithm $\mathcal{M}$ satisfies $(\alpha, \epsilon(\alpha))$-Rényi DP (RDP), for some $\alpha > 1$, if for any pair of adjacent datasets $X \sim X'$, $\mathsf{D}_\alpha(\mathbb{P}_{\mathcal{M}(X)} \| \mathbb{P}_{\mathcal{M}(X')}) \le \epsilon(\alpha)$. Here, $\mathbb{P}_{\mathcal{M}(X)}$ and $\mathbb{P}_{\mathcal{M}(X')}$ represent the distributions of $\mathcal{M}(X)$ and $\mathcal{M}(X')$, respectively, and*

$$\mathsf{D}_\alpha(\mathsf{P} \| \mathsf{Q}) = \frac{1}{\alpha - 1} \log \int \mathsf{q}(y) \left(\frac{\mathsf{p}(y)}{\mathsf{q}(y)}\right)^\alpha dy, \tag{19}$$

represents α-Rényi Divergence between two distributions P *and* Q *whose density functions are* p *and* q*, respectively.*

When the output domain $\mathcal{Y}^*$ of $\mathcal{M}$ is discrete, one can simply replace the integral in (19) by summation over elements $y \in \mathcal{Y}^*$ to obtain the discrete RDP version. RDP can be used to elegantly handle the composition of privacy leakage and enables a simple conversion to (ϵ, δ)-DP, as characterized below.

Proposition 3 (RDP Composition and Conversion to (ϵ, δ) DP [32])**.** *For any $\alpha > 1$, the class of $(\alpha, \epsilon_0(\alpha))$-RDP mechanisms satisfies (ϵ, δ)-differential privacy under T-fold adaptive composition for any ϵ and δ such that*

$$\epsilon \ge T\epsilon_0(\alpha) - \log(\delta)/(\alpha - 1). \tag{20}$$

Unfortunately, RDP cannot be directly applied to handle positive noise mechanisms. As demonstrated in Section 3, with one-sided or bounded noise, the support domains of the output distributions produced by two adjacent datasets, $\mathbb{P}_{\mathcal{M}(X)}$ and $\mathbb{P}_{\mathcal{M}(X')}$, *cannot* be exactly the same: there always exists some y within the support set of $\mathcal{M}(X)$ but beyond the support set of $\mathcal{M}(X')$ such that $\mathbb{P}\big(\mathcal{M}(X) = y\big) \neq 0$ while $\mathbb{P}\big(\mathcal{M}(X') = y\big) = 0$, which leads to an unbounded ratio $\mathbb{P}(\mathcal{M}(X) = y)/\mathbb{P}(\mathcal{M}(X') = y) = \infty$ for two adjacent datasets X and X'. Thus, the α Rényi divergence

$$\mathbb{E}_{\mathcal{M}(X')}\Big(\frac{\mathbb{P}(\mathcal{M}(X))}{\mathbb{P}(\mathcal{M}(X'))}\Big)^\alpha = \int_{-\infty}^{\infty} \frac{\big(\mathbb{P}(\mathcal{M}(X) = y)\big)^\alpha}{\big(\mathbb{P}(\mathcal{M}(X') = y)\big)^{\alpha-1}} dy \tag{21}$$

is not well-defined and Proposition 3 is not applicable. To this end, we present a generalization by computing the composition under positive noise in a hybrid form. The high-level idea is to measure the likelihood divergences separately in two cases: the part over the common (overlapped) support set is still measured through Rényi divergence while the remainder is controlled by a failure rate.

We consider an arbitrary processing function perturbed by some one-sided noise, denoted by $\mathcal{M} : \mathcal{X}^* \to \mathcal{Y}^*$. For an input set X, we use $S_d(X) \subset \mathcal{Y}^*$ to denote the subset of all degenerate events:

$$S_d(X) = \big\{y \mid \mathbb{P}(\mathcal{M}(X) = y) = 0\big\}. \tag{22}$$

Accordingly, we define *Partial $\big(\alpha, \mathcal{R}_{\alpha,p}(X, X')\big)$-Rényi Divergence* (PRD) for two adjacent datasets $X \sim X'$ as follows,

$$\mathcal{R}_{\alpha,p}(X, X') = \frac{1}{\alpha - 1} \log \int_{y \in \mathcal{Y}^*/S_d(X')} \frac{\mathbb{P}(\mathcal{M}(X) = y)^\alpha}{\mathbb{P}(\mathcal{M}(X') = y)^{\alpha-1}} dy. \tag{23}$$

Comparing (23) and (19), PRD only measures the divergence within the subset $\mathcal{Y}^*/S_d(X')$ where $\mathbb{P}(\mathcal{M}(X) = y) > 0$. In the following, we present Hybrid RDP (HRDP) to capture the cumulative privacy risk accounting for the release both within or outside the degenerate set S_d. We formally define $(\alpha, \epsilon_{\alpha,p}, \delta_p)$-HRDP as follows.

Definition 7 (Hybrid RDP). *A mechanism $\mathcal{M} : \mathcal{X}^* \to \mathcal{Y}^* \subset \mathbb{R}^d$ satisfies $(\alpha, \epsilon_{\alpha,p}, \delta_p)$-HRDP if for arbitrary two adjacent datasets X and X', the degenerate events are bounded as*

$$\sup_{X,X'} \Pr\big(\mathcal{M}(X') \in S_d(X)\big) \le \delta_p;$$

and their PRD defined in (23) is also bounded as

$$\sup_{X,X'} \mathcal{R}_{\alpha,p}(X, X') \le \epsilon_{\alpha,p}.$$

Theorem 2 (HRDP Composition). *For T mechanisms $\mathcal{M}_i, i = 1, 2, \cdots, T$ where each $\mathcal{M}_i$ satisfies $(\alpha, \epsilon^{(i)}_{\alpha,p}, \delta^{(i)}_p)$-HRDP, the composition of $\mathcal{M}_{[1:T]}$ satisfies (ϵ, δ)-DP such that for any $\delta' > 0$,*

$$\epsilon \geq \sum_{i=1}^{T} \epsilon^{(i)}_{\alpha,p} + \frac{\log(1/\delta')}{\alpha - 1}, \text{ and } \delta \geq \sum_{i=1}^{T} \delta^{(i)}_p + \delta'. \tag{24}$$

Compared to the regular composition of RDP described in Proposition 3, Theorem 2 demonstrates the following generalization: one can still apply partial Rényi divergence $\mathcal{R}_{\alpha,p}(X, X')$ between $\mathbb{P}_{\mathcal{M}(X)}$ and $\mathbb{P}_{\mathcal{M}(X')}$ over the non-degenerate domain $\mathcal{Y}^*/S_d(X')$, captured by $\epsilon^{(i)}_{\alpha,p}$, to bound ϵ in (24); on the other hand, the probability over the degenerate set $S_d(X')$, captured by $\delta^{(i)}_p$, is simply additive to the global failure rate δ.

For the discrete case, one can simply replace the integral in (23) by a summation and Theorem 2 still holds. As a final remark, the idea in Theorem 2 can be generalized to other composition accounting methods, for example, through the characteristic function [53], where we can similarly analyze the moment function over degenerate and non-degenerate domains, respectively.

4.3 Noise Optimization with Composition

With HRDP, now, we have a more powerful tool and a clearer characterization to handle the DP composition of general noise mechanisms. Given the objective global ϵ and δ bound in (24), respectively, a natural idea to determine the optimal noise distribution is to minimize (24) given the budget on δ, second moment B and maximal magnitude restriction R. However, one remaining obstacle here is that we need to specify the degenerate set S_d in the first place: the objective (24) varies with different selections of S_d. To address this, the following lemma provides a clearer picture on the concentration of the optimal positive noise distribution.

Lemma 2 (Contiguous Support Set). *To achieve (ϵ, δ)-DP under T-fold composition, the optimal bounded positive noise $e \in [0, R]$ with the minimal second moment must satisfy the following property:* $\Pr(e = 0) > 0$ *and if there exists some u such that* $\Pr(e = u) = 0$, *then* $\Pr(e \geq u) = 0$.

Lemma 2 states the following fact: the probability mass of optimal positive noise must be consecutively assigned along some interval $[0, R_0]$ for $R_0 \leq R$. To be specific, for a given R_0, let $\mathsf{P}_{R_0} = \{p_0, p_1, \cdots p_{R_0}\}$ denote a noise distribution e supported over $\{1, 2, \cdots, R_0\}$, where $\Pr(e = i) = p_i$. When sensitivity equals 1, in this discrete setup, the output distributions produced by two adjacent datasets X and X' are either identical or differ by a ± 1 shift, and without loss of generality, we assume $S_d(X) = \{p_{R_0}\}$ and $S_d(X') = \{p_0\}$. Therefore, by Lemma 2, the bound of ϵ in (24) can be equivalently expressed as a function $H(R_0, \mathsf{P}_{R_0})$ of R_0 and P_{R_0}:

$$H(R_0, \mathsf{P}_{R_0}) = \frac{1}{\alpha - 1} \max\{T \log \sum_{i=1}^{R_0} \frac{(p_i)^\alpha}{(p_{i-1})^{\alpha-1}} + \log(\frac{1}{\delta - Tp_0}), T \log \sum_{i=1}^{R_0} \frac{(p_{i-1})^\alpha}{(p_i)^{\alpha-1}} + \log(\frac{1}{\delta - Tp_{R_0}})\}.$$

$H(R_0, \mathsf{P}_{R_0})$ captures the worst case of HRDP by taking the maximal of $\epsilon_{\alpha,p}(X, X')$ with p_0 and $\epsilon_{\alpha,p}(X', X)$ with p_{R_0}. Therefore, we transform determining the optimal noise distribution into the following constrained optimization,

$$\min_{R_0} \min_{\mathsf{P}_{R_0}} \quad H(R_0, \mathsf{P}_{R_0}), \tag{25}$$

$$\text{s.t.} \quad \sum_{i=1}^{R_0} p_i = 1, 0 < p_i < 1, i = 1, 2, \cdots, R_0, \tag{26}$$

$$0 < R_0 \leq R, \sum_{i=0}^{R_0} i^2 \cdot p_i \leq B^2, p_0 < \frac{\delta}{T}, p_{R_0} < \frac{\delta}{T}. \tag{27}$$

In (26) and (27), we describe the constraints: (26) ensures that P_{R_0} is a distribution; (27) ensures that e is supported within $[0, R]$ and the second moment of P_{R_0} is bounded by B^2, and the probability of degenerate events S_d is bounded by δ/T. However, (25) is not directly solvable since as the selection of range R_0 varies, both the form of objective $H(R_0, \mathsf{P}_{R_0})$ and the constraints (26) and (27) change, due to a different p_{R_0} in S_d. To address this, we consider decomposing the original optimization over $H(R_0, \mathsf{P}_{R_0})$ into the optimization over the tail and R_0, respectively. We first consider the optimization on $H(R_0, \mathsf{P}_{R_0})$ restricted to noise distributions with fixed R_0, leftmost δ_l and rightmost δ_r tails. We introduce

$$\mathsf{D}(\delta_l, \delta_r, B, R_0) = \{\mathsf{P}_{R_0} : \sum_{i=0}^{R_0} i^2 p_i \leq B^2, p_0 = \delta_l, p_{R_0} = \delta_r\},$$

to capture the set of distributions supported on $[0 : R_0]$ with second moment bound B and fixed $p_0 = \delta_l$ and $p_{R_0} = \delta_r$.

Theorem 3 (Efficiency of Algorithm 1). *Given selections of δ_l, δ_r and R_0, minimization of $H(R_0, \mathsf{P}_{R_0})$ is equivalent to minimizing*

$$\max\{(\sum_{i=1}^{R_0} \frac{(p_i)^\alpha}{(p_{i-1})^{\alpha-1}}), (\sum_{i=1}^{R_0} \frac{(p_{i-1})^\alpha}{(p_i)^{\alpha-1}})\},$$

which is convex with respect to P_{R_0}. In addition, given R_0 and P_{R_0}, $H(R_0, \mathsf{P}_{R_0})$ is also convex with respect to p_0 and p_{R_0}, respectively.

The proof can be found in Appendix E. The above theorem demonstrates the following facts:

(a) Given δ_l, δ_r and R_0, minimizing $H(R_0, \mathsf{P}_{R_0})$ for $\mathsf{P}_{R_0} \in \mathsf{D}(\delta_l, \delta_r, B, R_0)$ is a convex optimization over a convex constraint set $\mathsf{D}(\delta_l, \delta_r, B, R_0)$.
(b) Given R_0 and P_{R_0}, $H(R_0, \mathsf{P}_{R_0})$ is convex w.r.t. the leftmost and rightmost slot p_0 and p_{R_0}.

To efficiently approximate the optimal noise distribution, we consider fixing $R_0 = R$ and utilizing the convexity shown in Theorem 3 to propose a two-layer algorithm to alternatively optimize (δ_l, δ_r) and P_R, as Algorithm 1.

4.4 Experiments and Comparisons

In this subsection, we set out to produce a set of experiments to show the power of both hybrid RDP (HRDP) accounting (Theorem 2) and optimized noise distribution (Algorithm 1). In practice, the second moment budget B and the maximal magnitude R of noise e capture the expected additional overhead and the worst-case redundancy required, respectively. For example, in anonymous communication [39, 41] or database operations [8, 14], B and R correspond to the expected and maximal dummy messages sent or files written. In Fig. 2(a), under $T = 500$ compositions, we show the required *second*

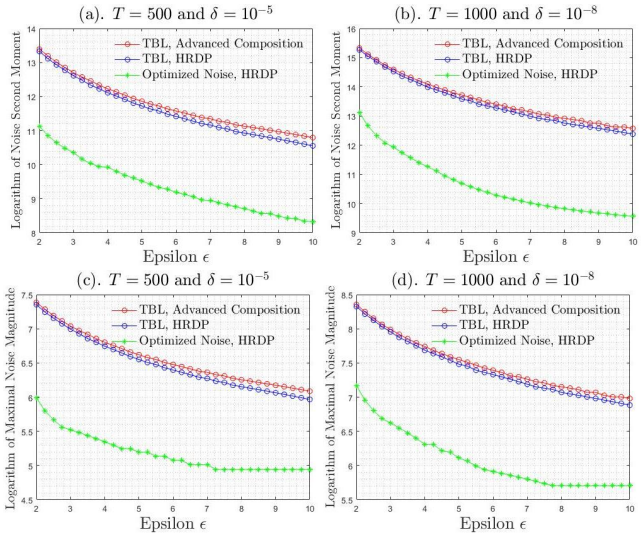


Figure 2: Second moment and maximal magnitude comparison between different positive DP noises and accounting.

.

moment $\mathbb{E}[e^2]$ of positive noise to produce (ϵ, δ)-DP with δ fixed to be 10^{-5} in the following three scenarios: 1) by discrete TBL noise in Definition 4 [8, 14, 41, 52] using advanced composition (Proposition 1); 2) still by discrete TBL noise, but using HRDP for composition accounting (Theorem 2); 3) the optimized noise based on HRDP and Algorithm 1. The (1,2,3) cases are captured by the red, blue and green lines, respectively. With a similar setup, in Fig. 2(b), we consider a scenario with more compositions $T = 1,000$ and also a smaller failure rate budget $\delta = 10^{-8}$. Comparing Case 1) (red line) and Case 2) (blue line), we can see HRDP produces a tighter composition bound. Additionally, when we compare Case 2) (blue line) and Case 3) (green line), it can be seen that after optimization, we significantly improve the noise to produce the *same* (ϵ, δ) parameter: the second moment of optimized noise is around 10× and 20× smaller than that of TBL in Fig. 2(a) and (b), respectively. Such improvement is *more significant* with a larger composition T and a smaller privacy budget (ϵ, δ).

With the same setup as Fig. 2(a,b), in Fig. 2(c,d), we consider the required *maximal magnitude* R for different noises with different accounting methods. Similarly, after optimization, one may find a better noise distribution accommodated in a larger interval with a weaker requirement on R. Compared to TBL noise, the maximal magnitude of optimized noise required is generally 5× smaller in these two examples. To give a more concrete example, we consider the same setup in the anonymous communication protocol *Vuvuzela* [41], where each user has a $T = 250,000$ message budget (composition) for a global budget $\epsilon = \log(2), \delta = 10^{-4}$. By TSL noise and advanced composition [41], the expected and the worst-case number of dummy messages sent per communication round is 300K and 600K, respectively, with an end-to-end additional latency around 18.4 seconds. As a comparison, by selecting a much smaller $R = 30,000$ in Algorithm 1, the optimized noise requires only 14K and 30K messages on average and in the worst case, respectively, with latency shortened to 1 second.

Algorithm 1 Optimized Positive Noise for Hybrid RDP

1: **Input:** Second moment budget B, range restriction R, composition number T, failure probability budget δ.
2: Randomly initialize both tail rate $\delta_l, \delta_r \in (0, \delta/T)$, and accordingly initialize $\mathsf{P}_R \in \mathsf{D}(\delta_l, \delta_r, B, R)$.
3: Alternately run convex optimizer on (δ_l, δ_r) with respect to the loss function $H(R, \mathsf{P}_R)$ with updated P_R, and on P_R with respect to the loss function $H(R, \mathsf{P}_R)$ given updated (δ_l, δ_r) such that $\delta_l, \delta_r \in (0, \delta/T)$, until they converge.
4: **Output**: P_R.

5 Positive Noise for Maximal Leakage

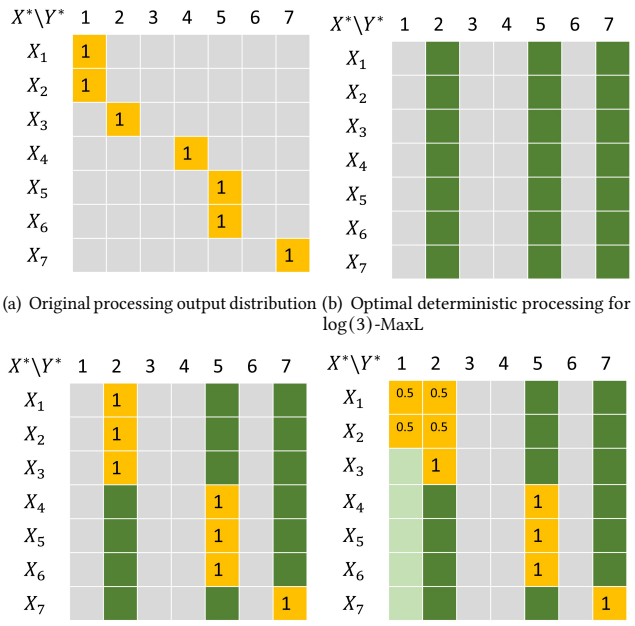
5.1 Theory and Algorithm

In this section, we study the optimal positive perturbation for the MaxL measure. We consider a deterministic function $\mathcal{F} : \mathcal{X}^* = \{X_1, X_2, \cdots, X_n\} \to \mathcal{Y}^* = \{1, 2, \cdots, m\}$, whose input domain $\mathcal{X}^*$ is formed by m possible selections and the output domain, capturing the information leakage, has n states. A state in $\mathcal{Y}^*$ of larger number (shortened to higher state in the following) has higher overhead, for example, taking a longer processing time, producing more cache misses or requiring a larger memory.

To randomize $\mathcal{F}$ with positive noise, we consider the randomized version of $\mathcal{F}$, denoted by $\mathcal{RF}$, where similarly $\mathcal{RF} : \mathcal{X}^* = \{X_1, X_2, \cdots, X_n\} \to \mathcal{Y}^* = \{1, 2, \cdots, m\}$. It is noted that an arbitrary (either deterministic or randomized) processing function can always be represented by a transition matrix $\mathsf{P}_{n\times m}$, where the entry at the crossing of the i-th row and j-th column (denoted as $p_{ij} \in [0, 1]$) represents the probability that we map X_i to the j-th state in the output domain $\mathcal{Y}^*$. The positive perturbation constraint requires that the support set of $\mathcal{RF}(X_i)$ is within $[\mathcal{F}(X_i), m]$.

To provide more intuition, we illustrate the corresponding transition matrix $\mathsf{P}_{n\times m}$ for the original processing function $\mathcal{F}$ and its positively perturbed version $\mathcal{RF}$, in Fig. 3(a) and Fig. 3(c), 3(d), respectively. We consider an example when $n = m = 7$, and in Fig. 3, each orange box represents a non-zero probability p_{ij} to map X_i to the j-th state. Adding positive noise is intuitively "decomposing" and "moving" the orange boxes in Fig. 3(a) to the right hand side. In Fig. 3(c) and 3(d), we give two positive perturbation schemes. For example, in Fig. 3(c), $\mathcal{F}(X_1) = 1$ (corresponding to $p_{11} = 1$), and for a feasible positive perturbation, $\mathcal{RF}$ is only allowed to map X_1 to higher states, starting from 1 to 7. The perturbed $\mathcal{RF}$ version in Fig. 3(c) changes the mapping of X_1 from state 1 to 2; in Fig. 3(d), the solution becomes that with 0.5 probability $\mathcal{RF}(X_1) = 1$ and with 0.5 probability $\mathcal{RF}(X_1) = 2$.

As defined in (4), the MaxL privacy leakage is measured as $\mathcal{L}(X \to Y) = \log \sum_{y\in\mathcal{Y}^*} \max_{x\in\mathcal{X}^*} \mathbb{P}(Y = y|X = x)$ which equals $\log\left(\sum_{j=1}^{m} \max_i p_{ij}\right)$. Given the transition matrix $\mathsf{P}_{n\times m}$, it can be viewed as the logarithm of the sum of the maximal element p_{ij} in each column. In the following, we formalize the cost to produce a mechanism with a satisfied MaxL loss. We introduce an $n \times m$ cost matrix $C_{n\times m}$ where each entry c_{ij} represents the cost of mapping X_i to the j-th output state. Under the positive noise restriction, $c_{ij} = \infty$ for $j < \mathcal{F}(X_i)$, i.e., the cost of mapping X_i to lower states compared to the original $\mathcal{F}(X_i)$ is formidably large. In addition, we also assume that for any fixed i, c_{ij} for $j \geq \mathcal{F}(X_i)$ is in



(a) Original processing output distribution (b) Optimal deterministic processing for log(3)-MaxL

(c) Water-filled perturbation of original processing for log(3)-MaxL (d) Water-filled perturbation of original processing for log(3.5)-MaxL

Figure 3: Illustration by transition matrix $\mathsf{P}_{n\times m}: \mathcal{X}^* \to \mathcal{Y}^*$ for positive noise perturbation in Maximal Leakage (MaxL)

a non-decreasing order, i.e., the cost to the higher states is more expensive. Putting the privacy risk and the cost together, determining the optimal perturbation translates to the following constrained optimization problem on transition matrix $\mathsf{P} = \{p_{ij}\}_{n\times m}$:

$$\min_{\mathsf{P}} C(\mathsf{P}) = \sum_{i=1}^{n} q_i \sum_{j=1}^{m} c_{ij} p_{ij} \text{ s.t. } \log\Big(\sum_{j=1}^{m} \max_i p_{ij}\Big) \le \log(v). \tag{28}$$

Here, q_i is the prior distribution of X_i, which captures the frequency that X_i is selected in the processing $\mathcal{F}$.

Before we introduce our main results to determine the optimum of (28), we need to point out a special case of $\mathcal{RF}$ with deterministic modifications to $\mathcal{F}$. In general, to ensure $\log(v)$-MaxL for an integer $v \ge 1$, a sufficient method is to simply select v states in $\mathcal{Y}^*$ and ensure that the support set of $\mathcal{RF}$ on any X_i is within them. An example is illustrated in Fig. 3(b), where we select $v = 3$ columns (states $\{2, 5, 7\}$) of the transition matrix $\mathsf{P}_{n\times m}$ and define an $\mathcal{RF}$ which maps X_i to the *closest* higher state $\{2, 5, 7\}$ compared to $\mathcal{F}(X_i)$ (see Fig. 3(b)). For example, $\mathcal{F}(X_1) = 1$ moves to the closest higher state selected $\mathcal{RF}(X_1) = 2$; $\mathcal{F}(X_4) = 4$ becomes $\mathcal{RF}(X_4) = 5$.

The idea of finding the closest higher state can be further generalized into a *water-filling* operation and the representation of the optimal processing scheme can be further simplified from a transition matrix $\mathsf{P}_{n\times m}$ to an m-dimensional vector $\bar{\mathcal{P}}_m$, as studied in [44]. Let $\bar{\mathcal{P}}_m = (\bar{p}_1, \bar{p}_2, \cdots, \bar{p}_m)$, where $\bar{p}_j$ represents the upper bound of p_{ij} in the j-th column. $\bar{\mathcal{P}}_m$ ensures $\log \sum_{j=1}^{m} \bar{p}_j$-MaxL. We say that $\bar{\mathcal{P}}_m$ dominates a transition matrix $\mathsf{P}_{n\times m}$ iff $p_{ij} \le \bar{p}_j$ for any i, j. Given a selection of $\bar{\mathcal{P}}_m$, for all transition matrices dominated by $\bar{\mathcal{P}}_m$, the one minimizing the cost function (28) must be in a *water-filling* form, i.e., starting from the state $\mathcal{F}(X_i)$, the optimal mechanism would iteratively fill the j-th slot up to probability $\bar{p}_j$ for $j = \mathcal{F}(X_i), \mathcal{F}(X_i)+1, \cdots, m$, until the sum $\sum_j p_{ij}$ becomes 1. A formal statement and conclusion is given as follows.

Proposition 4 (Water-filling Lemma [44]). *Among all mechanisms with transition matrix $\mathsf{P} = \{p_{ij} | i \in [1:n]; j \in [1:m]\}$ dominated by $\bar{\mathcal{P}}_m = (\bar{p}_1, \bar{p}_2, \cdots, \bar{p}_m)$, the optimal one with the minimal cost is in the following form: for any i and j, $p_{ij} = \min\{\bar{p}_j, 1 - \sum_{k=\mathcal{F}(X_i)}^{j-1} \bar{p}_k\}$.*

Thus, the $\bar{\mathcal{P}}_m$ for a deterministic mechanism must be a binary vector and in particular, if we restrict a feasible positively perturbed $\mathcal{RF}$ to be deterministic for v-MaxL for some integer $v \ge 1$, it is *equivalent* to selecting v states, denoted by $\mathcal{A} = \{a_1, a_2, \cdots, a_v\}$, between $\min_i \mathcal{F}(X_i)$ and $\max_i \mathcal{F}(X_i)$ and sets $\bar{p}_j = 1$ for $j \in \mathcal{A}$ and it is noted that $\max_i \mathcal{F}(X_i)$ must be selected in $\mathcal{A}$.. Therefore, determining the optimal deterministic solution for $\log(v)$-MaxL with an integer v is reduced to searching over all v-subsets of states with Proposition 4 and comparing their optimal water-filling schemes. A formal definition is given below.

Definition 8 (Optimal Deterministic Scheme for Integer v). *When $v \in \mathbb{Z}^+$, we define DS_v as the set of the deterministic scheme(s) that achieve the minimal cost conditioned on $\log(v)$-MaxL.*

With the above understanding, there are still two remaining challenges to fully characterize the optimal positive perturbation for MaxL. First and more fundamentally, even when v is integer, is the optimal deterministic scheme also the global optimum for all (possibly randomized) schemes, and for general v, is the optimal solution related to the optimal deterministic schemes in some way? Second, enumeration-based searching takes $O(n \cdot \binom{m}{v})$ time. Can we more efficiently determine the optimal scheme? We will answer both questions affirmatively in the following.

Theorem 4 (Optimal Perturbation for MaxL). *When v is some positive integer, the optimal solution(s) to (28) are exactly DS_v. When $v = \lceil v \rceil - \lambda$ for $\lambda \in (0, 1)$ is not an integer, then the optimal solution(s) to (28) is the linear interpolation of $\mathsf{DS}_{\lfloor v \rfloor}$ and $\mathsf{DS}_{\lceil v \rceil}$ as,*

$$\lambda \cdot \mathsf{DS}_{\lfloor v \rfloor} + (1-\lambda) \cdot \mathsf{DS}_{\lceil v \rceil} = \{\lambda \cdot \mathsf{P}_{\lfloor v \rfloor} + (1-\lambda) \cdot \mathsf{P}_{\lceil v \rceil} \mid \mathsf{P}_{\lfloor v \rfloor} \in \mathsf{DS}_{\lfloor v \rfloor}, \mathsf{P}_{\lceil v \rceil} \in \mathsf{DS}_{\lceil v \rceil}\}.$$

Theorem 4, with proof in Appendix F, is an improvement of the results in [44] and states the following facts. For $\log(v)$-MaxL with integer v, the optimal solution should be deterministic and in DS_v. For $\log(v)$-MaxL with non-integer v, both the cost function and the privacy function are linear with respect to the elements in $\mathsf{DS}_{\lfloor v \rfloor}$ and $\mathsf{DS}_{\lceil v \rceil}$, and the optimal solution is in a weighted average of two arbitrary deterministic solutions from $\mathsf{DS}_{\lfloor v \rfloor}$ and $\mathsf{DS}_{\lceil v \rceil}$, respectively. Theorem 4 also paves the path to efficiently determine the optimal solution, where it suffices to find optimal deterministic schemes in DS_v (or $\mathsf{DS}_{\lfloor v \rfloor}$ and $\mathsf{DS}_{\lceil v \rceil}$). This is solvable by dynamic programming in $O(n \cdot m^2)$ time.

To be specific, we introduce $\mathcal{T}(i, k)$ (sub-algorithm 2 in Algorithm 2) that considers the optimal deterministic mechanism when (1) $\bar{p}_1, \cdots, \bar{p}_{i-1}$ are given and $\bar{p}_{i-1} = 1$; and (2) $\sum_{j=i}^{m} \bar{p}_j$ equals k. This condition means that given the current selection of $\bar{p}_1, \cdots, \bar{p}_{i-1}$, we still need to pick k additional states within $[a:m]$. Note that there are totally $n \cdot m$ possible inputs for $\mathcal{T}(\cdot, \cdot)$. For any $\mathcal{T}(i, k)$ such that $k > 0$, we consider the next state to select

in the optimal scheme, i.e., the minimal $j \geq i$ such that $\bar{p}_j = 1$. Once j is given, $\mathcal{T}(i,k)$ is reduced to the sub-problem $\mathcal{T}(j,k-1)$. This means that $\mathcal{T}(i,k)$ can be solved once we solve $\mathcal{T}(j,k-1)$ for all $j \geq i$. Therefore, we can use dynamic programming to solve the problem and the time complexity is $O(n \cdot m^2)$, as detailed in Appendix G.

As an illustration, in Fig. 3(c) and 3(d), we show the optimal positively-perturbed scheme with minimal cost $\min_{\mathsf{P}} C(\mathsf{P})$ for $v = 3$ and $v = 3.5$ when we select the prior distribution $q_i = 1/n$ to be uniform and $c_{ij} = j$ if $j \geq \mathcal{F}(X_i)$, otherwise $c_{ij} = \infty$.

5.2 Applications: AES Secret Key Protection

We provide a concrete application of Algorithm 2 to determine the optimal perturbation scheme to control the MaxL of cache leakage when implementing AES in an S-box implementation [4]. We adopt the same setup and evaluation method as in *Metior* (more details can be found in Section 7 of [19]). Described in MaxL language, we consider $\mathcal{U}^*$ to be a 256-bit secret key space, i.e., $\mathcal{U}^* = \{0,1\}^{256}$. $\mathcal{X}^*$ corresponds to the intermediate number of distinct cache lines the program touches [12]. Finally, the observation $y \in \mathcal{Y}^*$ captures the number of misses. We use *Metior* [19] to determine both the prior distribution on $\mathcal{X}^*$ and the mapping from $\mathcal{X}^*$ to $\mathcal{Y}^*$. Based our evaluation in this example, $\mathcal{X}^*$ is formed by 370 intermediate states and $\mathcal{Y}^* = \{1, 2, \cdots, 38\}$, i.e., the maximal number of misses is bounded by 38 in all cases. In Fig. 4(a), we consider two kinds of cost functions $C_{n\times m}$. The black line corresponds to the case where $c_{ij} = j$ if $j \geq \mathcal{F}(X_i)$ otherwise $c_{ij} = \infty$, which takes the expected number of misses as the cost. The blue line captures the case where $c_{ij} = j - \mathcal{F}(X_i)$ if $j \geq \mathcal{F}(X_i)$ otherwise $c_{ij} = \infty$, which captures the expected number of *additional* misses. As illustrated by Fig. 4(a), for $\log(v)$-MaxL, the optimal cost $\min_{\mathsf{P}} C(\mathsf{P})$ is linear when $v \in [\lfloor v \rfloor, \lceil v \rceil]$. Moreover, as v increases, the cost decreases and approaches that of the original processing function. To further interpret the results, it is noted that the prior success rate for an adversary to correctly identify a 256-bit secret key is 2^{-256}. From Fig. 4 we show that at a minimal cost of 6.8 and 1 dummy misses on average, the adversarial posterior success rate to correctly recover the secret key after observing the leakage is bounded, negligibly, by 2^{-253} and 2^{-245}, respectively.

Finally, we want to emphasize that proposed Algorithm 2 is the *first* efficient algorithm to provably determine the optimal scheme in MaxL; prior works either only measure the MaxL of some heuristic protocols without providing privatization solutions [19] or only approximate the optimal perturbation [44].

6 Positive Noise for PAC Privacy

In this section, we study how to determine the optimal postive noise for PAC Privacy. From Proposition 2, we know the posterior advantage measured in KL-divergence $\Delta^{\rho}_{KL} = \mathcal{D}_{KL}(\mathbf{1}_{\delta_\rho} \| \mathbf{1}_{\delta_{o,\rho}})$, for an arbitrary adversarial inference captured by ρ, is upper bounded by the mutual information $\mathsf{MI}(X; \mathcal{F}(X)+\mathsf{e})$, where X is the sensitive input generated from some distribution D, $\mathcal{F}$ is the processing function and e is some positive noise. As one of the key motivations of PAC Privacy is to enable automatic privatization of black-box processing $\mathcal{F}$, here, we do not put any additional assumptions on the output distribution $Y = \mathcal{F}(X)$ but only assume the variance of Y, $\mathrm{Var}(Y) = \mathbb{E}\big[(Y - \mathbb{E}[Y])^2\big]$, is bounded by σ_Y^2. As demonstrated in [46], when the output domain $\mathcal{Y}^*$ of $\mathcal{F}$ is bounded, $\mathrm{Var}(Y)$ can be estimated efficiently in high confidence. Within this general black-box setup, we do *not* assume either the input $\mathcal{X}^*$ or output domain $\mathcal{Y}^*$ of the processing function $\mathcal{F} : \mathcal{X}^* \to \mathcal{Y}^*$ is finite, and $\mathcal{F}(X)$ for $X \leftarrow \mathrm{D}$ can be either discrete or continuous.

In the following, we focus on positive noise e in a continuous distribution within some bounded interval $[0, R]$. Still, the utility loss is defined by the second moment of e, $\mathbb{E}[\mathsf{e}^2] = \mu_e^2 + \sigma_e^2$, required to be bounded by some budget B. Here, we use μ_e and σ_e^2 to denote

Algorithm 2 Optimal Mechanism for Maximal Leakage

Input: Objective processing function $\mathcal{F} : \mathcal{X}^* = \{X_1, X_2, \cdots, X_N\} \to \{1, 2, \cdots, m\}$, prior distribution $\mathbb{P}_{\mathcal{X}^*}$ of input over $\mathcal{X}^*$ where $p_i = \Pr(X = X_i)$; cost weight c_{ij} of mapping X_i to the state j; objective MaxL budget $\log(v)$.
if v is integer **then**
  Run Sub-algorithm 1 to determine the optimal deterministic mechanism $\mathcal{M}_{\mathcal{D}}(v)$ and output $\mathcal{M}_{\mathcal{D}}(v)$.
else
  Run Sub-algorithm 1 to determine the respective optimal deterministic mechanisms $\mathcal{M}_{\mathcal{D}}(\lfloor v \rfloor)$ and $\mathcal{M}_{\mathcal{D}}(\lceil v \rceil)$.
  Let $\lambda = \lceil v \rceil - v$.
  Output $\lambda \mathcal{M}_{\mathcal{D}}(\lfloor v \rfloor) + (1-\lambda)\mathcal{M}_{\mathcal{D}}(\lceil v \rceil)$.
end if

Sub-algorithm 1: Optimal Deterministic Mechanism $\mathcal{M}_{\mathcal{D}}$. Takes as input an integer $k = v$ and returns the optimal deterministic mechanism in vector form.

if $k = 1$ **then**
  Returns $(0, 0, \cdots, 0, 1)$, which allocates all input to m.
else
  Initialize $i \leftarrow 0$ and $p \leftarrow (0, 0, \cdots, 0)$.
  for k' in order of $k, k-1, \cdots, 1$ **do**
    $i \leftarrow \mathcal{T}(i, k').Next$.
    $p_i \leftarrow 1$.
  end for
  Return $p = (p_1, \cdots, p_m)$.
end if

Sub-algorithm 2: Dynamic Programming algorithm $\mathcal{T}(a,k)$. $\mathcal{T}$ takes as inputs a position $a \in [m]$ and the remaining budget k.

if $k = 1$ **then**
  $Next \leftarrow m$.
  $Cost \leftarrow \sum_{(i \mid \mathcal{F}(X_i) \geq a)} p_i \cdot c_{i,m}$.
  Return $(Next, Cost)$.
else if $a \geq m + 1$ **then**
  Return $(null, 0)$.
else
  for a' in $\{a+1, , \cdots, m+1\}$ **do**
    $cost_{a'} \leftarrow \mathcal{T}(a', k-1).Cost + \sum_{a \leq X_i \leq a'-1} p_i \cdot c_{i,(a'-1)}$.
  end for
  $Next \leftarrow \arg\min_{a'} cost_{a'}$.
  $Cost \leftarrow \min_{a'} cost_{a'}$.
  Return $(Next, Cost)$.
end if

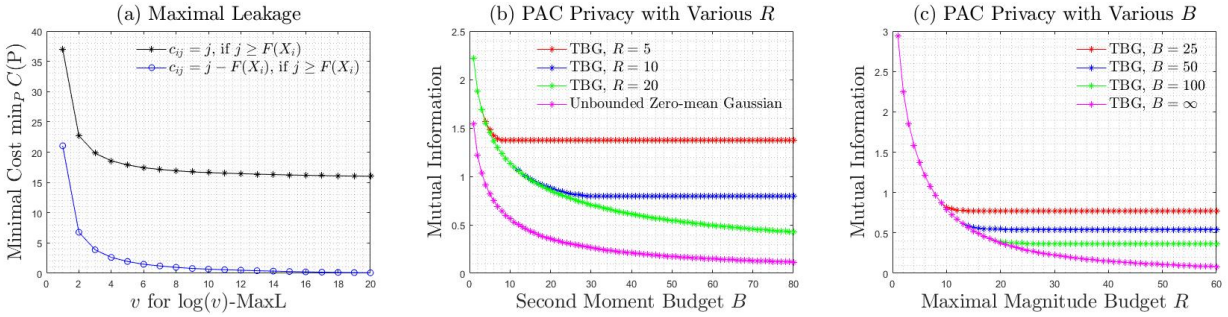


Figure 4: AES Secret Key Protection with Optimal Positive Perturbation in Maximal Leakage and PAC Privacy

.

the mean (bias) and the variance of the noise e, respectively. To determine the optimal noise distribution D_e, it reduces to solve the following min-max problem,

$$\begin{aligned} &\min_{\mathsf{D}_e, \mathbb{E}[e^2] \leq B} \quad \max_{\mathsf{D}_{\mathcal{F}(X)}, \mathrm{Var}(\mathcal{F}(X)) \leq \sigma_Y^2} \mathsf{MI}(X; \mathcal{F}(X) + e) \\ &= \min_{\mathsf{D}_e, \mathbb{E}[e^2] \leq B} \quad \max_{\mathsf{D}_Y, \mathrm{Var}(Y) \leq \sigma_Y^2} \mathsf{h}(Y + e) - \mathsf{h}(e). \end{aligned} \tag{29}$$

In (29), we use h to represent the differential entropy, where $\mathsf{h}(e) = \int_{-\infty}^{\infty} -\mathbb{P}(e = z) \log(\mathbb{P}(e = z))dz$, and adopt the classic entropy expression of mutual information [17]. (29) models that given a second moment budget B of injected noise e, we aim to minimize the worst-case mutual information for arbitrary processing output distribution $Y = \mathcal{F}(X)$ with bounded variance σ_Y^2, the only knowledge assumed about the black-box processing function $\mathcal{F}$.

On the other hand, since Y and e are independent, the variance of $Y + e$ is upper bounded by $\sigma_Y^2 + \sigma_e^2$. It is well-known that for any continuous distribution with bounded variance, the Gaussian distribution achieves the maximal entropy, where $h(Y + e) \leq \frac{1}{2} \cdot \big(\log(2\pi(\sigma_Y^2 + \sigma_e^2)) + 1\big)$[17]. The equality is achievable and we specify it in Appendix H. Thus, (29) is upper bounded as

$$\min_{\mathsf{D}_e, \mathbb{E}[e^2] \leq B} \frac{1}{2} \cdot \big(\log(2\pi(\sigma_Y^2 + \sigma_e^2)) + 1\big) - \mathsf{h}(e). \tag{30}$$

The following theorem demonstrates that the optimal distribution to (30) under constraints must be a Gaussian distribution $\mathcal{N}(\mu, \sigma)$ truncated and normalized within $[0, R]$.

Theorem 5 (Optimal Positive Noise Class for PAC Privacy). *Suppose e is restricted within $[0, R]$ and $\mathbb{E}[e^2] \leq B$, then the optimal noise distribution D_e of (30) must be in a truncated Gaussian form*

$$\mathbb{P}(e = z) = \frac{1}{\sigma(\Phi(\frac{R-\mu}{\sigma}) - \Phi(\frac{-\mu}{\sigma}))} \cdot e^{-(z-\mu)^2/(2\sigma^2)} \cdot \mathbf{1}_{z \in [0,R]},$$

for some $\mu \in [0, R]$, i.e., some Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ conditional on $[0, R]$. Here, $\Phi(t)$ is the cumulative probability function of $\mathcal{N}(0, 1)$, i.e., $\Phi(t) = \int_{-\infty}^{t} 1/\sqrt{2\pi} \cdot e^{-t^2/2} dt$.

The proof of Theorem 5 can be found in Appendix H. By Theorem 5 with proof in Appendix H, it suffices to optimize the two parameters μ and σ of a truncated Gaussian distribution within $[0, R]$. Since both σ_e and $\mathsf{h}(e)$ in (30) can be expressed using μ and σ, the objective problem with the associated second moment constraint can be rewritten as

Algorithm 3 Generating noise for PAC privacy

Input: The maximum range for the positive noise R; the second moment budget B; the standard deviation of output y, σ_y.
Use binary search to find σ_o that minimizes $\mathcal{L}_{std}(\sigma)$.
Return $\mathcal{L}_{std}(\sigma_o)$.

Sub-algorithm $\mathcal{L}_{std}$: given a standard deviation σ, find the best mean μ such that the normal distribution $N(\mu, \sigma)$ gives the optimal PAC privacy loss. Return the corresponding PAC loss.

if $\mathcal{S}(\frac{R}{2}, \sigma) \leq B$ **then**
  Return $\frac{R}{2}$.
else
  Use binary search to find μ' such that $\mathcal{S}(\mu', \sigma) = B$ and return μ'.
end if

In (31), we use the following notations: $\alpha = -\mu/\sigma$ and $\beta = (R - \mu)/\sigma$ with $\varphi(x) = 1/\sqrt{2\pi} \cdot e^{-x^2/2}$ and $\Phi(t) = \int_{-\infty}^{t} 1/\sqrt{2\pi} \cdot e^{-t^2/2} dt$.

At a first glance, both the objective function $\mathrm{obj}(\mu, \sigma)$ and the second moment constraint $\mathcal{S}(\mu, \sigma) \leq B$ in (31) are complicated without a closed form. However, assisted by symbolic analysis in *Mathematica*, we have the following important observations:

a) Given σ, if we ignore the second-moment budget B (or equivalently set $B = \infty$), $\mathrm{Obj}(\mu, \sigma)$ in (31) decreases when $\mu < R/2$ and increases when $\mu > R/2$, and the minimum is achieved when $\mu = R/2$;

b) Given σ, the second moment $\mathcal{S}(\mu, \sigma)$ of a Gaussian noise $\mathcal{N}(\mu, \sigma^2)$ of mean μ and variance σ^2 truncated over $[0, R]$ strictly increases with μ.

The above two observations suggest that when σ is given, we can adjust the μ value to reduce $\mathrm{Obj}(\mu, \sigma)$ as follows:

(1) We first evaluate $\mathcal{S}(R/2, \sigma)$. If $\mathcal{S}(R/2, \sigma) \leq B$, then return $R/2$ as the optimal μ;
(2) Otherwise, we need to find the maximum μ such that $\mathcal{S}(\mu, \sigma) \leq B$. Given that $\mathcal{S}(\mu, \sigma)$ increases with μ, this is equivalent to finding μ such that $\mathcal{S}(\mu, \sigma) = B$ and we can use binary search to find an approximation up to any κ accuracy within $\Theta(\log(1/\kappa))$ time.

Now, let $\mu(\sigma)$ denote the optimal μ given σ, and provided a) and b), $\mathrm{Obj}(\mu(\sigma), \sigma)$ also enjoys a nice property, which first decreases and then increases with σ, leading to a unique optimum. This allows a

$$\min_{\mu,\sigma} \ \mathrm{Obj}(\mu,\sigma) = \frac{1}{2}\cdot\log\left(\sigma_Y^2+\sigma^2\Big[1-\frac{\beta\varphi(\beta)-\alpha\varphi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}-\big(\frac{\varphi(\beta)-\varphi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}\big)^2\Big]\right)-\log\Big(\sqrt{2\pi e}\cdot\sigma(\Phi(\beta)-\Phi(\alpha))\Big)-\frac{\alpha\varphi(\alpha)-\beta\varphi(\beta)}{2(\Phi(\beta)-\Phi(\alpha))},$$
$$\text{such that} \qquad \mathcal{S}(\mu,\sigma)=\Big(\mu-\sigma\cdot\frac{\varphi(\beta)-\varphi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}\Big)^2+\sigma^2\Big[1-\frac{\beta\varphi(\beta)-\alpha\varphi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}-\big(\frac{\varphi(\beta)-\varphi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}\big)^2\Big]\le B. \tag{31}$$

straightforward use of binary search again to find a good σ value, summarized as Algorithm 3.

With Algorithm 3, we continue the study on cache leakage of an AES key in Section 5.2 and present near-optimal positive noise in PAC privacy in Fig. 4 (b,c). In the same setup, following the same evaluation by Metior [19], the variance of released cache misses from a random 256-bit secret key equals $\sigma_Y^2 = 20.6$. In Figure 4(b), we plot the leakage measured in mutual information across various second moment budget B when $R = 5, 10, 20$. Note that given R, the μ of optimal noise must satisfy $\mu \le \frac{R}{2}$. This is because the selections of (μ,σ) and $(R-\mu,\sigma)$ produce the same privacy loss, but $(R-\mu,\sigma)$ has a larger second moment. Thus, the second moment of the optimal noise is bounded by $R^2/3$ and the privacy loss will decrease with B until $B = R^2/3$. This matches the observation from Figure 4(b), where all curves of TBG noise become flat after some turning points. Meanwhile, there is a gap between the privacy loss produced by positive TBG noise and unbounded, zero-mean Gaussian noise [46], which narrows as R and B increase.

In Fig. 4(c), we plot the privacy loss against the maximum magnitude R, when $B = 25, 50, 100$, and ∞, respectively. Here, we also observe that the privacy loss stops decreasing with large enough R. Recall that if R is large enough such that $\mathcal{S}(R/2,\sigma) > B$, the optimal μ should be selected such that $\mathcal{S}(\mu,\sigma) = B$. Therefore, after R reaches the threshold determined by $\mathcal{S}(R/2,\sigma) = B$, further increasing R only results in minor changes to the μ and σ of the optimal distribution, and thus provides limited improvement to the privacy risk bound.

7 Additional Related Works

Bounded Noise Mechanism: From an asymptotic viewpoint, the maximal magnitude R of perturbation to produce DP guarantees under compositions has been studied in [18]. In particular, [18] proved that when $\delta \ge e^{-\tilde{O}(T)}$, there exists some bounded noise within $[-R,R]$ such that $R = O(\sqrt{T\log(1/\delta)}/\epsilon)$ to produce (ϵ,δ)-DP under T-fold composition. This suggests that when δ is not too small, the bounded noise mechanism can asymptotically match the same second moment of a general noise mechanism such as Gaussian without constraints. However, though the noise construction of [18] is shown to enjoy the asymptotically-optimal dependence on R, non-asymptotically, its second moment may not be always optimal and can be worse than TSL in the examples considered in Fig. 2. As a comparison, in this paper, we show how to non-asymptotically optimize the noise distribution given various budgets with the more powerful HRDP to handle the composition.

Optimal Mechanism for MaxL: The properties of MaxL and its relationship to other information theory quantities, such as Sibson mutual information, have been studied in [28, 36] and the optimal scheme for MaxL has also been previously studied in [28]. In particular, [28] proved that, with proper non-decreasing assumptions on the cost matrix $C_{n\times m}$, the convex hull of the cost functions determined by deterministic schemes is identical to that formed by both deterministic and randomized schemes. In this paper, we improve their results and show a stronger linear interpolation representation to fully characterize the optimal scheme for arbitrary $log(v)$-MaxL using the optimal deterministic schemes in Theorem 4. Based on Theorem 4, we present the first efficient algorithm to determine the optimal solution in polynomial time.

8 Conclusions and Future Work

In this paper, we studied the optimal positive perturbation in various privacy metrics. With a focus on the utility loss measured by *(weighted) second moment* of noise, we provided insights into the characteristics of the optimal perturbation using finite parameters that enable efficient optimization. However, it is worth noting that the expected absolute value of the injected noise may not be the most accurate measure of the resulting utility loss. It would be interesting to generalize our techniques to handle more general utility loss measurements.

We also want to mention, besides mitigating side-channel leakage, one-sided noise is also useful for applications with a specific privacy-preserving overestimation (optimistic) or underestimation (pessimistic) requirement. In addition, for our results on DP, it should be noted that Rényi DP is not the tightest known method to compute composition; even more powerful tools are known, such as, f-DP [20] or through characteristic functions [53]. Thus, one interesting generalization is to consider the hybrid version of those more advanced composition accounting methods. One may follow a similar idea to consider the hybrid versions of those measures on positive noise mechanisms.

For PAC Privacy, in this paper we focused on the general black-box processing without assuming anything specific about the distribution of $\mathcal{F}(X)$ except its variance. When more information is given with respect to $\mathcal{F}$, an interesting problem is to generalize Theorem 5 to accommodate the stronger prior knowledge on $\mathcal{F}$.

References

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. 308–318.

[2] Alejandro Cabrera Aldaya, Cesar Pereida García, Luis Manuel Alvarez Tapia, and Billy Bob Brumley. 2018. Cache-timing attacks on RSA key generation. *Cryptology ePrint Archive* (2018).

[3] Mário S Alvim, Konstantinos Chatzikokolakis, Annabelle McIver, Carroll Morgan, Catuscia Palamidessi, and Geoffrey Smith. 2020. *The science of quantitative information flow*. Springer.

[4] Paulo Barreto Vincent Rijmen and Antoon Bosselaers. [n. d.]. Optimised ANSI C code for the Rijndael cipher. https://opensource.apple.com/source/BerkeleyDB/BerkeleyDB-15/db/crypto/rijndael/rijndael-alg-fst.c.auto.html

[5] Raef Bassily, Adam Smith, and Abhradeep Thakurta. 2014. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *2014 IEEE 55th annual symposium on foundations of computer science*. IEEE, 464–473.

[6] Johes Bater, Xi He, William Ehrich, Ashwin Machanavajjhala, and Jennie Rogers. 2018. Shrinkwrap: efficient sql query processing in differentially private data federations. *Proceedings of the VLDB Endowment* 12, 3 (2018).

[7] Daniel J Bernstein. [n. d.]. Cache-timing attacks on AES. ([n. d.]).

[8] Dmytro Bogatov, Georgios Kellaris, George Kollios, Kobbi Nissim, and Adam O'Neill. 2021. εpsolute: Efficiently Querying Databases While Providing Differential Privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. 2262–2276.

[9] Christelle Braun, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. 2009. Quantitative notions of leakage for one-try attacks. *Electronic Notes in Theoretical Computer Science* 249 (2009), 75–91.

[10] Mark Bun and Thomas Steinke. 2016. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*. Springer, 635–658.

[11] Benedikt Bünz, Yuncong Hu, Shin'ichiro Matsuo, and Elaine Shi. 2021. Non-interactive differentially anonymous router. *Cryptology ePrint Archive* (2021).

[12] Cristian Cadar, Daniel Dunbar, Dawson R Engler, et al. 2008. Klee: Unassisted and automatic generation of high-coverage tests for complex systems programs.. In *OSDI*, Vol. 8. 209–224.

[13] Benjamin M Case, James Honaker, and Mahnush Movahedi. 2021. The Privacy-preserving Padding Problem: Non-negative Mechanisms for Conservative Answers with Differential Privacy. *arXiv preprint arXiv:2110.08177* (2021).

[14] T-H Hubert Chan, Kai-Min Chung, Bruce Maggs, and Elaine Shi. 2022. Foundations of differentially oblivious algorithms. *ACM Journal of the ACM (JACM)* 69, 4 (2022), 1–49.

[15] T-H Hubert Chan, Kai-Min Chung, Bruce Maggs, and Elaine Shi. 2022. Foundations of differentially oblivious algorithms. *ACM Journal of the ACM (JACM)* 69, 4 (2022), 1–49.

[16] Keith Conrad. 2004. Probability distributions and maximum entropy. *Entropy* 6, 452 (2004), 10.

[17] Thomas M Cover. 1999. *Elements of information theory*. John Wiley & Sons.

[18] Yuval Dagan and Gil Kur. 2022. A bounded-noise mechanism for differential privacy. In *Conference on Learning Theory*. PMLR, 625–661.

[19] Peter W Deutsch, Weon Taek Na, Thomas Bourgeat, Joel S Emer, and Mengjia Yan. 2023. Metior: A Comprehensive Model to Evaluate Obfuscating Side-Channel Defense Schemes. In *Proceedings of the 50th Annual International Symposium on Computer Architecture*. 1–16.

[20] Jinshuo Dong, Aaron Roth, and Weijie J Su. 2022. Gaussian differential privacy. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84, 1 (2022), 3–37.

[21] Cynthia Dwork. 2006. Differential privacy. In *International colloquium on automata, languages, and programming*. Springer, 1–12.

[22] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. 2006. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology-EUROCRYPT 2006: 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28-June 1, 2006. Proceedings 25*. Springer, 486–503.

[23] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*. Springer, 265–284.

[24] Cynthia Dwork, Aaron Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9, 3–4 (2014), 211–407.

[25] Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. 2010. Boosting and differential privacy. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. IEEE, 51–60.

[26] Marco Gaboardi, Michael Hay, and Salil Vadhan. 2020. A programming framework for opendp. *Manuscript, May* (2020).

[27] Quan Geng and Pramod Viswanath. 2015. The optimal noise-adding mechanism in differential privacy. *IEEE Transactions on Information Theory* 62, 2 (2015), 925–951.

[28] Ibrahim Issa, Aaron B Wagner, and Sudeep Kamath. 2019. An operational approach to information leakage. *IEEE Transactions on Information Theory* 66, 3 (2019), 1625–1657.

[29] Paul C Kocher. 1996. Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems. In *Advances in Cryptology—CRYPTO'96: 16th Annual International Cryptology Conference Santa Barbara, California, USA August 18–22, 1996 Proceedings 16*. Springer, 104–113.

[30] Albert Kwon, Henry Corrigan-Gibbs, Srinivas Devadas, and Bryan Ford. 2017. Atom: Horizontally scaling strong anonymity. In *Proceedings of the 26th Symposium on Operating Systems Principles*. 406–422.

[31] David Lazar and Nickolai Zeldovich. 2016. Alpenhorn: Bootstrapping secure communication without leaking metadata. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. 571–586.

[32] Ilya Mironov. 2017. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*. IEEE, 263–275.

[33] S Alvim M'rio, Kostas Chatzikokolakis, Catuscia Palamidessi, and Geoffrey Smith. 2012. Measuring information leakage using generalized gain functions. In *2012 IEEE 25th Computer Security Foundations Symposium*. IEEE, 265–279.

[34] Lianke Qin, Rajesh Jayaram, Elaine Shi, Zhao Song, Danyang Zhuo, and Shumo Chu. 2022. Adore Differentially Oblivious Relational Database Operators. *Proceedings of the VLDB Endowment* 16, 4 (2022), 842–855.

[35] Lianke Qin, Rajesh Jayaram, Elaine Shi, Zhao Song, Danyang Zhuo, and Shumo Chu. 2022. Adore: Differentially Oblivious Relational Database Operators. *Proceedings of the VLDB Endowment* 16, 4 (2022), 842–855.

[36] Sara Saeidian, Giulia Cervia, Tobias J Oechtering, and Mikael Skoglund. 2023. Pointwise maximal leakage. *IEEE Transactions on Information Theory* (2023).

[37] Igal Sason and Sergio Verdú. 2016. f-divergence Inequalities. *IEEE Transactions on Information Theory* 62, 11 (2016), 5973–6006.

[38] Mayuri Sridhar, Hanshen Xiao, and Srinivas Devadas. 2025. PAC-Private Algorithms. In *2025 IEEE Symposium on Security and Privacy (SP)*. IEEE.

[39] Nirvan Tyagi, Yossi Gilad, Derek Leung, Matei Zaharia, and Nickolai Zeldovich. 2017. Stadium: A distributed metadata-private messaging system. In *Proceedings of the 26th Symposium on Operating Systems Principles*. 423–440.

[40] Jo Van Bulck, Marina Minkin, Ofir Weisse, Daniel Genkin, Baris Kasikci, Frank Piessens, Mark Silberstein, Thomas F Wenisch, Yuval Yarom, and Raoul Strackx. 2018. Foreshadow: Extracting the keys to the intel {SGX} kingdom with transient {Out-of-Order} execution. In *27th USENIX Security Symposium (USENIX Security 18)*. 991–1008.

[41] Jelle Van Den Hooff, David Lazar, Matei Zaharia, and Nickolai Zeldovich. 2015. Vuvuzela: Scalable private messaging resistant to traffic analysis. In *Proceedings of the 25th Symposium on Operating Systems Principles*. 137–152.

[42] Tim Van Erven and Peter Harremos. 2014. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory* 60, 7 (2014), 3797–3820.

[43] Di Wang, Hanshen Xiao, Srinivas Devadas, and Jinhui Xu. 2020. On differentially private stochastic convex optimization with heavy-tailed data. In *International Conference on Machine Learning*. PMLR, 10081–10091.

[44] Benjamin Wu, Aaron B Wagner, and G Edward Suh. 2020. Optimal mechanisms under maximal leakage. In *2020 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 1–6.

[45] Hanshen Xiao. 2024. *Automated and Provable Privatization for Black-Box Processing*. Ph. D. Dissertation. Massachusetts Institute of Technology.

[46] Hanshen Xiao and Srinivas Devadas. 2023. PAC Privacy: Automatic Privacy Measurement and Control of Data Processing. In *Annual International Cryptology Conference*. Springer, 611–644.

[47] Hanshen Xiao, G Edward Suh, and Srinivas Devadas. 2024. Formal Privacy Proof of Data Encoding: The Possibility and Impossibility of Learnable Obfuscation. In *Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security*.

[48] Hanshen Xiao, Jun Wan, and Srinivas Devadas. 2023. Geometry of sensitivity: twice sampling and hybrid clipping in differential privacy with optimal gaussian noise and application to deep learning. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*. 2636–2650.

[49] Hanshen Xiao, Zihang Xiang, Di Wang, and Srinivas Devadas. 2023. A theory to instruct differentially-private learning via clipping bias reduction. In *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2170–2189.

[50] Xiaokui Xiao and Yufei Tao. 2008. Output perturbation with query relaxation. *Proceedings of the VLDB Endowment* 1, 1 (2008), 857–869.

[51] Ashkan Yousefpour, Igor Shilov, Alexandre Sablayrolles, Davide Testuggine, Karthik Prasad, Mani Malek, John Nguyen, Sayan Ghosh, Akash Bharadwaj, Jessica Zhao, et al. 2021. Opacus: User-friendly differential privacy library in PyTorch. *arXiv preprint arXiv:2109.12298* (2021).

[52] Mingxun Zhou, Elaine Shi, T-H Hubert Chan, and Shir Maimon. 2023. A theory of composition for differential obliviousness. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 3–34.

[53] Yuqing Zhu, Jinshuo Dong, and Yu-Xiang Wang. 2022. Optimal accounting of differential privacy via characteristic function. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 4782–4817.

A Proof of Lemma 1

Lemma 1 (Parameter of Positive Laplace Noise). *Suppose a processing function $\mathcal{F}: \mathcal{X}^* \to \mathbb{R}$ such that for an arbitrary adjacent dataset pair $X \sim X'$, $|\mathcal{F}(X) - \mathcal{F}(X')| \le \mathrm{s}$, i.e., the sensitivity of $\mathcal{F}$ is bounded by* s. *Then, if we select $\lambda_L = \mathrm{s}/\epsilon$, $\mu_L \ge \mathrm{s} + \frac{\mathrm{s}}{\epsilon} \cdot \log \frac{1}{2\delta(1-e^{-\mu_L \cdot \epsilon/\mathrm{s}})}$, and $R = 2\mu_L$, such a (μ_L, λ_L, R)-TBL perturbation ensures (ϵ, δ)-DP.*

Proof. Without loss of generality, given the (ϵ, δ) measure is invariant to the shift, we assume $\mathcal{F}(X) = 0$ and $\mathcal{F}(X') = \mathrm{s}$. Thus, the support domain of the distribution of $\mathcal{F}(X) + \mathrm{e}$ is $[0, R]$ while that of $\mathcal{F}(X) + \mathrm{e}$ is over $[\mathrm{s}, R + \mathrm{s}]$.

First, given the selection of μ_L and λ_L with $R = 2\mu_L$, once $\mu_L \ge \mathrm{s}$, it is noted that

$$Z_{\mu,\lambda,R} = \frac{1}{1 - e^{-\mu_L/\lambda_L}},$$

and

$$\Pr\left(\mathrm{e} \in [0, \mathrm{s}]\right) = \frac{0.5\left(e^{-(\mu_L - \mathrm{s})/\lambda_L} - e^{-\mu_L/\lambda_L}\right)}{1 - e^{-\mu_L/\lambda_L}}.$$

Thus, we need to ensure $\Pr\left(\mathrm{e} \in [0, \mathrm{s}]\right) \le \delta$ (and similarly $\Pr\left(\mathrm{e} \in [R - \mathrm{s}, R]\right) \le \delta$), which in turn suffices to show

$$e^{-(\mu_L - \mathrm{s})/\lambda_L} \le e^{-(\mu_L - \mathrm{s})/\lambda_L} - e^{-\mu_L/\lambda_L} \le 2\delta(1 - e^{-\mu_L/\lambda_L}),$$

which suffices to require that

$$\frac{\mu_L - \mathrm{s}}{\lambda_L} \ge \log \frac{1}{2\delta(1 - e^{-\mu_L/\lambda_L})}.$$

On the other hand, based on the property of Laplace noise, conditional on the output of $\mathcal{F}(X) + e$ is within $[0, R - \mathrm{s}]$, where $\mathcal{F}(X') + e$ is within $[\mathrm{s}, R]$, when $\lambda = \mathrm{s}/\epsilon$, it is not hard to see that the distribution of $\mathcal{F}(X) + \mathrm{e}$ and that of $\mathcal{F}(X') + \mathrm{e}$ satisfies the divergence requirement for (ϵ, δ)-DP.

□

B Proof of Theorem 1

Theorem 1 (Optimum for Single Release). *Given a processing function $\mathcal{F}$ of sensitivity 1, among all possible distributions of a positive noise* e *over $[0, +\infty)$ which ensure an (ϵ, δ)-DP guarantee of the noisy version $\mathcal{F}(\cdot) + \mathrm{e}$, the following distribution with probability mass function given in (1) below,*

$$p_i = \begin{cases} \delta \cdot e^{\epsilon i} & \text{if } i < \omega \\ \delta \cdot c \cdot e^{\epsilon(2\omega - i)} & \text{if } \omega \le i \le \omega', \end{cases} \tag{16}$$

is optimal in a sense that it achieves the minimal k-th moment, for any *positive integer k. Here, ω' is either $2\omega - 1$ or 2ω, and $c \in [e^{-2\varepsilon}, 1]$ is for normalization such that the sum of p_i equals 1. Here, ω is a turning point, defined as*

$$\omega = \frac{1}{\epsilon} \cdot \log\left(\frac{2}{e^{\epsilon} + 1} + \frac{e^{\epsilon} - 1}{\delta(e^{\epsilon} + 1)}\right). \tag{17}$$

To prove Theorem 1, we start with two lemmas as follows.

Lemma 3. *If a distribution $(p_0, p_1, \cdots)$ satisfies (ϵ, δ)-DP under sensitivity 1 and $p_i > \delta$, then for any $j \le \ln(p_i/\delta)/\epsilon$, $p_{i+j} \ge e^{-j\epsilon} \cdot p_i$.*

Proof. Denote $L = \{i \mid p_i > e^{\epsilon} p_{i+1}\}$. If $\{p_i\}$ satisfies (ϵ, δ)-DP, then $\sum_{i \in R} p_i \le \delta$.

We will prove the lemma by induction. The lemma clearly holds when $j = 0$. Suppose the lemma holds for some $j \le \ln(p_i/\delta)/\epsilon - 1$ as well, i.e.,

$$p_{i+j} \ge e^{-j\epsilon} p_i.$$

This implies that $p_{i+j} > \delta$, thus $(i + j)$ can not be in L. Therefore, $p_{i+j+1} \ge e^{-\epsilon} p_{i+j}$ and the lemma holds for $j + 1$ as well. □

Lemma 4. *Given two tuples $P = (p_0, p_1, \cdots, p_n)$ and $Q = (q_0, q_1, \cdots, q_n)$ where $\sum_{i=0}^{n} p_i = \sum_{i=0}^{n} q_i$, if $\sum_{j=0}^{i} p_j \ge \sum_{j=1}^{i} p_j$ holds for all $0 \le i \le n$, then for any $v_0 \le v_1 \le \cdots \le v_n$, we have $\sum_{i=0}^{n} p_i v_i \le \sum_{i=0}^{n} q_i v_i$.*

Proof. We can prove by induction on n. First, we know that

$$p_n - q_n = \sum_{i=0}^{n-1} q_i - \sum_{i=0}^{n-1} p_i \le 0.$$

We will construct a new tuple $Q' = (q'_0, \cdots, q'_{n-1}, q'_n)$ where (1) $q'_n = p_n$, (2) $q'_{n-1} = q_{n-1} + (q_n - p_n)$ and (3) $q'_i = q_i$ for i in $\{0, 1, \cdots, n-2\}$. Essentially, we reduce Q's weight in the last position to p_n and move the reduced amount to position $n - 1$. We have

$$\sum_{i=0}^{n} q_i v_i = (q_n - p_n) \cdot (v_n - v_{n-1}) + \sum_{i=0}^{n} q'_i v_i \ge \sum_{i=0}^{n} q'_i v_i.$$

Notice that $\sum_{j=0}^{i} p_j \ge \sum_{j=0}^{i} q'_j$ still holds between P and Q'. Further, $p_n = q'_n$. By induction, the lemma should hold for the tuples $(p_0, \cdots, p_{n-1})$ and $(q'_0, \cdots, q'_{n-1})$. Thus,

$$\sum_{i=0}^{n} q_i v_i \ge \sum_{i=0}^{n} q'_i v_i \ge \sum_{i=0}^{n} p_i v_i.$$

This completes our proof. □

We will now prove Theorem 1.

Proof. We first specify how the parameters ω' and c are chosen. Recall that in Theorem 1, we define

$$\omega = \frac{1}{\epsilon} \cdot \log\left(\frac{2}{e^{\epsilon} + 1} + \frac{e^{\epsilon} - 1}{\delta(e^{\epsilon} + 1)}\right),$$

Let

$$c(\omega) = \frac{e^{\epsilon} - 1 - (e^{\epsilon\omega} - 1) \cdot \delta}{(e^{\epsilon(\omega+1)} - 1) \cdot \delta}.$$

If $c(\omega) \ge e^{-2\epsilon}$, then we set $\omega' = 2\omega$ and $c = c(\omega)$. Else, we set $\omega' = 2\omega - 1$ and

$$c = \frac{e^{\epsilon} - 1 - (e^{\epsilon\omega} - 1) \cdot \delta}{(e^{\epsilon(\omega+1)} - e^{\epsilon}) \cdot \delta}.$$

It can be guaranteed that $c \ge e^{-2\epsilon}$ and $p_i \ge \delta$ except for $i = \omega'$.

Let $(p_0, p_1, \cdots)$ be the distribution constructed in (16) using the above parameters. We first show that it satisfies (ε, δ)-DP. By definition, for any $i \in [0, \omega' - 1]$,

$$\frac{p_i}{p_{i+1}} = \begin{cases} e^{-\epsilon} & \text{if } i < \omega - 1 \\ e^{-\epsilon}/c & \text{if } i = \omega - 1 \\ e^{\epsilon} & \text{if } i > \omega - 1. \end{cases}$$

Since $c \in [e^{-2\epsilon}, 1]$, we have $(p_i/p_{i+1}) \in [e^{-\epsilon}, e^{\epsilon}]$ for all $i \in [0, \omega' - 1]$. Therefore, the only two points that violate ϵ-DP are $i = 0$ and

$i = \omega'$. Given that $p_0 = \delta$ and $p_{\omega'} \le \delta$, the distribution $(p_0, p_1, \cdots)$ satisfies (ϵ, δ)-DP.

Next, we show that the noise in (16) is optimal. For any other distribution $(p'_0, p'_1, \cdots)$ that satisfies (ϵ, δ), we claim that for all $i \ge 0$,

$$\sum_{j=0}^{i} p_j \ge \sum_{j=0}^{i} p'_j. \tag{32}$$

This means that p is a strictly "smaller" distribution than p'. For any $m > 0$, if we set $v_i = i^m$ in Lemma 4, then we have that for any $m > 0$, $\sum_i p_i \cdot i^m \le \sum_i p'_i \cdot i^m$, which means that p has a smaller m^{th} moment than p'. Let

$$L = \{i \mid p'_i > e^\epsilon p'_{i+1}\} \text{ and } R = \{i \mid p'_i > e^\epsilon p'_{i-1}\}.$$

By definition of (ϵ, δ)-DP, we have

$$\Pr[i \in L] = \sum_{i \in L} p'_i \le \delta \text{ and } \Pr[i \in R] = \sum_{i \in R} p'_i \le \delta.$$

Since p'_{-1} is undefined, $0 \in R$ and it must be that $p'_0 \le \delta = p_0$. Therefore, (32) holds when $i = 0$.

We first prove using induction that for any $1 \le i < \omega$, $p'_i \le p_i$. Suppose this holds for some $i = k - 1$, let us consider when $i = k$.

- If $k \in R$, then $p'_k \le \delta \le p_k$.
- If $k \notin R$, then $p'_k \le e^\epsilon p'_{k-1} \le e^\epsilon p_{k-1} = p_k$.

Since $p'_i \le p_i$ for all $1 \le i \le \omega$, this immediately implies that (32) holds for all $1 \le i \le \omega$.

We now consider when $\omega \le i < \omega'$. Suppose (32) holds for $i = k - 1$ ($\omega \le k < \omega'$), and let us consider the scenario when $i = k$. Let us assume that (32) does not hold for $i = k$, i.e.,

$$\sum_{j=0}^{k} p_j < \sum_{j=0}^{k} p'_j. \tag{33}$$

Since (32) holds for $i = k - 1$, it must be that $p_k < p'_k$. By Lemma 3, for any $j \le \omega' - k$,

$$p'_{k+j} \ge e^{-j\epsilon} \cdot p'_k > e^{-j\epsilon} \cdot p_k = p_{k+j}$$

holds for all $k + j \le 2T$. Therefore,

$$\sum_{i=0}^{\omega'} p'_i = \sum_{i=0}^{k} p'_i + \sum_{i=k+1}^{\omega'} p'_i > \sum_{i=0}^{k} p_i + \sum_{i=k+1}^{\omega'} p_i = 1.$$

The first part is based on our assumption in (33) and the second part is based on $p'_i > p_i$ for all $k \le i \le \omega'$. We have reached a contradiction. Thus the lemma must also hold for $i = k$.

Finally, note that the lemma trivially holds for $i = \omega'$, as $\sum_{i=0}^{\omega'} p'_i \le 1 = \sum_{i=0}^{\omega'} p_i$. This completes our induction proof. □

C Proof of Theorem 2

Theorem 2 (HRDP Composition). *For T mechanisms $\mathcal{M}_i, i = 1, 2, \cdots, T$ where each $\mathcal{M}_i$ satisfies $(\alpha, \epsilon^{(i)}_{\alpha,p}, \delta^{(i)}_p)$-HRDP, the composition of $\mathcal{M}_{[1:T]}$ satisfies (ϵ, δ)-DP such that for any $\delta' > 0$,*

$$\epsilon \ge \sum_{i=1}^{T} \epsilon^{(i)}_{\alpha,p} + \frac{\log(1/\delta')}{\alpha - 1}, \text{ and } \delta \ge \sum_{i=1}^{T} \delta^{(i)}_p + \delta'. \tag{24}$$

We will use the following lemma mildly generalized from Proposition 10 in [32].

Lemma 5. *Let $\alpha > 1$, P and Q be two distributions defined over $\mathbb{R}$, with probability density function p and q, respectively. Let $S_d(\mathrm{Q}) = \{z | q(z) = 0\}$ be the degenerate set of Q. Then, for any $A \subset \mathbb{R}/S_d(\mathrm{Q})$,*

$$\mathrm{P}(A) \le \left(e^{\mathrm{D}_\alpha(\mathrm{P}\cdot\mathbf{1}_A \| \mathrm{Q}\cdot\mathbf{1}_A)} \cdot \mathrm{Q}(A)\right)^{(\alpha-1)/\alpha}. \tag{34}$$

Proof. Based on the Holder Inequality,

$$\begin{aligned} \mathrm{P}(A) &= \int_{z\in A} p(z)dz \\ &\le \Big(\int_{z\in A} p(z)^\alpha q(z)^{1-\alpha} dz\Big)^{1/\alpha} \cdot \Big(\int_{z\in A} q(z)dz\Big)^{(\alpha-1)/\alpha} \\ &= \left(e^{\mathrm{D}_\alpha(\mathrm{P}\cdot\mathbf{1}_A \| \mathrm{Q}\cdot\mathbf{1}_A)} \cdot \mathrm{Q}(A)\right)^{(\alpha-1)/\alpha}. \end{aligned} \tag{35}$$

□

Now, we consider the composition. Let $Y_1, Y_2, \cdots, Y_T$ be the independent outputs of the objective mechanism $\mathcal{M}^{\otimes T}$ across T iterations. For two arbitrary adjacent datasets X and X', suppose $\delta_0(X) = \Pr(\mathcal{M}^{\otimes T}(X) \in S_d(X'))$. By union bound, the probability

$$\Pr(\mathcal{M}^{\otimes T}(X) = (Y_1, Y_2, \cdots, Y_T) \in (\mathbb{R}/S_d(X'))^{\otimes T}) \ge 1 - T\delta_0.$$

On the other hand, let

$$\epsilon_0(X) = \mathrm{D}_\alpha(\mathbb{P}_{\mathcal{M}(X)} \cdot \mathbf{1}_{\mathbb{R}/S_d(X')} \| \mathbb{P}_{\mathcal{M}(X')} \mathbf{1}_{\mathbb{R}/S_d(X')})$$

be the partial RDP conditioned on the set $\mathbb{R}/S_d(X')$. Now, for an arbitrary set $A \in \mathbb{R}^T$, let $A_{nd} = A \cap (\mathbb{R}/S_d(X'))^{\otimes T}$, and we have that

$$\begin{aligned} &\Pr(\mathcal{M}^{\otimes T}(X) \in A) \\ &= \Pr(\mathcal{M}^{\otimes T}(X) \in A_{nd}) + \Pr(\mathcal{M}^{\otimes T}(X) \in A/A_{nd}) \\ &\le T\delta_0(X) \\ &\quad + \left(e^{\mathrm{D}_\alpha(\mathbb{P}^{\otimes T}_{\mathcal{M}(X)} \cdot \mathbf{1}_{A_{nd}} \| \mathbb{P}^{\otimes T}_{\mathcal{M}(X')} \mathbf{1}_{A_{nd}})} \Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd})\right)^{\frac{\alpha-1}{\alpha}} \\ &= T\delta_0(X) + \left(e^{T\epsilon_0(X)} \Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd})\right)^{(\alpha-1)/\alpha}. \end{aligned} \tag{36}$$

In the last equation in (36), we use the fact that the Rényi Divergence between independent product distributions equals the sum of Rényi Divergences between each corresponding pair. Now, with a similar reasoning as Proposition 3 in [32], for some $\delta' > 0$, if $\left(e^{T\epsilon_0(X)} \Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd})\right) > (\delta')^{\alpha/(\alpha-1)}$, then

$$\begin{aligned} &\left(e^{T\epsilon_0(X)} \Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd})\right)^{1-1/\alpha} \\ &\le \left(e^{T\epsilon_0(X)} \Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd})\right) \cdot (\delta')^{-1/(\alpha-1)} \\ &= \left(e^{T\epsilon_0(X) + \frac{\log(1/\delta')}{\alpha-1}}\right) \cdot \Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd}). \end{aligned} \tag{37}$$

For the other case when $\left(e^{T\epsilon_0(X)} \Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd})\right) \le (\delta')^{\alpha/(\alpha-1)}$, it is clear that

$$\left(e^{T\epsilon_0(X)} \Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd})\right)^{1-1/\alpha} \le \delta'.$$

Therefore, putting things together, we obtain that

$$\begin{aligned} \Pr(\mathcal{M}^{\otimes T}(X) \in A) &\le (T\delta_0(X) + \delta') \\ &\quad + \left(e^{T\epsilon_0(X) + \frac{\log(1/\delta')}{\alpha-1}}\right) \cdot \Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd}), \end{aligned} \tag{38}$$

which provides the expression of ϵ and δ in (24), respectively.

D Proof of Lemma 2

Lemma 2 (Contiguous Support Set). *To achieve (ϵ, δ)-DP under T-fold composition, the optimal bounded positive noise $e \in [0, R]$ with the minimal second moment must satisfy the following property:* $\Pr(e = 0) > 0$ *and if there exists some u such that* $\Pr(e = u) = 0$, *then* $\Pr(e \geq u) = 0$.

Proof. Suppose a noise distribution $(p_0, p_1, \cdots,)$ satisfies (ϵ, δ)-DP under 1-sensitivity and there exists some $k \geq 0$ such that (1) $p_{k-1} \neq 0$, (2) $p_k = 0$, and (4) $\sum_{i>k} p_i \neq 0$. We will show that the following distribution

$$p'_i = \begin{cases} p_i & \text{if } i < k \\ p_{i+1} & \text{if } i \geq k, \end{cases}$$

also satisfies (ϵ, δ)-DP. Similar to the proof of Theorem 1, let us define

$$L(p) = \{i \mid p_i > e^\epsilon p_{i+1}\} \text{ and } R(p) = \{i \mid p_i > e^\epsilon p_{i-1}\}.$$

The key observation is that

$$\begin{cases} i \in L(p) \iff i \in L(p') & \text{if } i < k-1 \\ i \in L(p) & \text{if } i = k-1 \\ i \in L(p) \impliedby (i-1) \in L(p') & \text{if } i > k-1. \end{cases}$$

It immediately follows that $\Pr[i \in L(p)|p] \leq \Pr[i \in L(p')|p']$. A formal analysis is provided as follows:

$$\begin{aligned} &\Pr[i \in L(p)|p] \\ &= p_{i-1} + \Big(\sum_{i<k-1 \ \&\ i\in L(p)} p_i\Big) + \Big(\sum_{i>k-1 \ \&\ i\in L(p)} p_i\Big) \\ &\geq \Big(\sum_{i<k-1 \ \&\ i\in L(p')} p_i\Big) + \Big(\sum_{i>k-1 \ \&\ i-1\in L(p')} p_i\Big) \\ &= \Big(\sum_{i<k-1 \ \&\ i\in L(p')} p'_i\Big) + \Big(\sum_{i>k-1 \ \&\ i-1\in L(p')} p'_{i-1}\Big) \\ &= \Pr[i \in L(p')|p']. \end{aligned}$$

Similarly, it can be shown that $\Pr[i \in R(p)|p] \geq \Pr[i \in R(p')|p']$. This means that the probability that p violates ϵ-DP is higher than the probability that p' violates ϵ-DP. Therefore, p' satisfies (ϵ, δ)-DP as well.

Note that the second moment of p' is strictly smaller than that of p. So p cannot be the optimal noise. This concludes our proof for Lemma 2. □

E Proof of Theorem 3

Theorem 3 (Efficiency of Algorithm 1). *Given selections of δ_l, δ_r and R_0, minimization of $H(R_0, \mathsf{P}_{R_0})$ is equivalent to minimizing*

$$\max\Big\{\Big(\sum_{i=1}^{R_0} \frac{(p_i)^\alpha}{(p_{i-1})^{\alpha-1}}\Big), \Big(\sum_{i=1}^{R_0} \frac{(p_{i-1})^\alpha}{(p_i)^{\alpha-1}}\Big)\Big\},$$

which is convex with respect to P_{R_0}. In addition, given R_0 and P_{R_0}, $H(R_0, \mathsf{P}_{R_0})$ is also convex with respect to p_0 and p_{R_0}, respectively.

Proof. For convenience, let us denote $n = R_0$ in the following proof. We will show that the following function

$$H(p_0, \cdots, p_n) = \frac{1}{\delta - Tp_0} \cdot \Big(\sum_{i=1}^{n} \frac{(p_i)^\alpha}{(p_{i-1})^{\alpha-1}}\Big)^T$$

is convex on both p_0 and p_n, i.e.,

$$\frac{\partial^2 H(p_0, \cdots, p_n)}{\partial p_0^2} \geq 0, \text{ and } \frac{\partial^2 H(p_0, \cdots, p_n)}{\partial p_n^2} \geq 0.$$

Note that this does not imply that $H(p_0, \cdots, p_n)$ is convex on the space spanned by (p_0, p_n). Since we are focusing on p_0 and p_n, we can consider $p_1, \cdots, p_{n-1}$ as constants. Our observation is that $H(p_0, \cdots, p_n)$ can be written as a sum of sub-functions of the following form:

$$h(p_0, p_n) = \frac{c \cdot p_n^a}{(1 - p_0)p_0^b},$$

where $c > 0$ and $a, b \geq 1$ are some positive constants. If we can show that

$$\frac{\partial^2 h(p_0, p_n)}{\partial p_0^2} \geq 0 \text{ and } \frac{\partial^2 h(p_0, p_n)}{\partial p_n^2} \geq 0$$

hold as long as $c > 0$ and $a, b \geq 1$, then it naturally follows that $H(p_0, \cdots, p_n)$ is convex on both p_0 and p_n.

We consider the function

$$h(x, y) = \frac{y^a}{(1 - x)x^b},$$

where $a, b \geq 1$. The second partial derivative of $h(x, y)$ on x is

$$\frac{\partial^2 h(x, y)}{\partial x^2} = y^a \cdot \frac{x^{2b-2} \cdot ((b+1)(b+2)x^2 - 2bx + b^2)}{(x^b - x^{b+1})^3}.$$

Since $b \geq 1$, we have

$$\begin{aligned} &(b+1)(b+2)x^2 - 2bx + b^2 \\ &\geq b^2x^2 - 2bx + b^2 \\ &\geq b(x^2 - 2x + 1) \\ &\geq 0. \end{aligned}$$

Therefore,

$$\frac{\partial^2 h(x, y)}{\partial x^2} \geq 0.$$

The second partial derivative of $h(x, y)$ on y is

$$\frac{\partial^2 h(x, y)}{\partial y^2} = \frac{a(a-1)y^{a-2}}{(1 - x)x^b} \geq 0.$$

In the following, we prove the second part of this theorem. It is noted that once R_0, $p_0 = \delta_l$ and $p_{R_0} = \delta_r$ are given, minimizing

$H(R_0, \mathsf{P}_{R_0})$ with respect to P_{R_0} becomes

$$\arg\min_{\mathsf{P}_{R_0}} H(R_0, \mathsf{P}_{R_0}) \tag{39}$$

$$= \arg\min_{\mathsf{P}_{R_0}} \frac{1}{\alpha-1} \max\{T \log \sum_{i=1}^{R_0} \frac{(p_i)^\alpha}{(p_{i-1})^{\alpha-1}} + \log(\frac{1}{\delta - Tp_l}), \tag{40}$$

$$T \log \sum_{i=1}^{R_0} \frac{(p_{i-1})^\alpha}{(p_i)^{\alpha-1}} + \log(\frac{1}{\delta - Tp_r})\} \tag{41}$$

$$= \arg\min_{\mathsf{P}_{R_0}} \max\{\sum_{i=1}^{R_0} \frac{(p_i)^\alpha}{(p_{i-1})^{\alpha-1}}, \sum_{i=1}^{R_0} \frac{(p_{i-1})^\alpha}{(p_i)^{\alpha-1}}\}, \tag{42}$$

by removing the constant term and using the monotone property of $\log(\cdot)$. By the joint convexity of the Hellinger integral [42], it is known that for any two pairs of positive real numbers (p_0, q_0) and (p_1, q_1), and some $\lambda \in (0,1)$,

$$(1-\lambda)p_0^\alpha q_0^{1-\alpha} + \lambda p_1^\alpha q_1^{1-\alpha} \geq p_\lambda^\alpha q_\lambda^{1-\alpha}. \tag{43}$$

Here, $p_\lambda = (1-\lambda)p_0 + \lambda p_1$ and $q_\lambda = (1-\lambda)q_0 + \lambda q_1$. Therefore, both $\sum_{i=1}^{R_0} \frac{(p_i)^\alpha}{(p_{i-1})^{\alpha-1}}$ and $\sum_{i=1}^{R_0} \frac{(p_{i-1})^\alpha}{(p_i)^{\alpha-1}}$ in (42) are convex with respect to the distribution P_{R_0}, and it is not hard to verify that the max of two convex functions is still convex.

□

F Proof of Theorem 4

Theorem 4 (Optimal Perturbation for MaxL)**.** *When v is some positive integer, the optimal solution(s) to (28) are exactly* DS_v*. When $v = \lceil v \rceil - \lambda$ for $\lambda \in (0,1)$ is not an integer, then the optimal solution(s) to (28) is the linear interpolation of* $\mathsf{DS}_{\lfloor v \rfloor}$ *and* $\mathsf{DS}_{\lceil v \rceil}$ *as,*

$$\lambda \cdot \mathsf{DS}_{\lfloor v \rfloor} + (1-\lambda) \cdot \mathsf{DS}_{\lceil v \rceil}$$
$$= \{\lambda \cdot \mathsf{P}_{\lfloor v \rfloor} + (1-\lambda) \cdot \mathsf{P}_{\lceil v \rceil} \mid \mathsf{P}_{\lfloor v \rfloor} \in \mathsf{DS}_{\lfloor v \rfloor}, \mathsf{P}_{\lceil v \rceil} \in \mathsf{DS}_{\lceil v \rceil}\}.$$

Before we dive into Theorem 4, we need to first recap some of the results in [44]. Given a mechanism $\mathsf{P} = \{p_{ij}\}$ and some prior distribution $\{q_i\}$, we use $\mathcal{L}(\mathsf{P})$ to denote the exponential of the privacy loss and $C(\mathsf{P})$ to denote the cost, i.e.,

$$\mathcal{L}(\mathsf{P}) = \sum_{j=1}^{m} \max_i p_{ij}, \; C(\mathsf{P}) = \sum_{i=1}^{n} q_i \sum_{j=1}^{m} c_{ij} p_{ij}.$$

Note that the maximum leakage privacy loss is actually $\log(\mathcal{L}(\mathsf{P}))$. We will ignore the logarithmic function and focus on exploring the relationship between $\mathcal{L}(\mathsf{P})$ and $C(\mathsf{P})$. In this way, the loss $\mathcal{L}(\mathsf{P})$ is integer for any deterministic P. This simplifies our analysis. Let S be the set of achievable $(\mathcal{L}(\mathsf{P}), C(\mathsf{P}))$ pairs for any mechanism P, and let S_d be the set of achievable $(\mathcal{L}(\mathsf{P}), C(\mathsf{P}))$ pairs for any *deterministic* mechanism P. It is obvious that S_d is a subset of S. Further, if a mechanism P is optimal under its loss budget $\mathcal{L}(\mathsf{P})$, then $C(\mathsf{P}) = \inf[C : (\mathcal{L}(\mathsf{P}), C) \in S]$. It is shown in [44] that the boundary of the convex hull formed by S and S_d are the same.

Lemma 6. *If the cost function satisfies the requirement in Section 5.1, then for any $\alpha > 0$*

$$\min_{(L,C)\in S)} L + \alpha \cdot C = \min_{(L,C)\in S_d)} L + \alpha \cdot C.$$

In the rest of this appendix section, we will need to use both the vector and matrix representation of a mechanism. But when we discuss a linear combination of two mechanisms, by default, we are always talking about a linear combination in the vector representation. Specifically, given two mechanisms $\mathsf{P} = (\bar{p}_1, \cdots, \bar{p}_m)$ and $\mathsf{P}' = (\bar{p}'_1, \cdots, \bar{p}'_m)$, for any $\lambda \in [0,1]$, we define $\lambda\mathsf{P} + (1-\lambda)\mathsf{P}' = (\lambda\bar{p}_1 + (1-\lambda)\bar{p}'_1, \cdots, \lambda\bar{p}_m + (1-\lambda)\bar{p}'_m)$. Note that under vector representation $\mathsf{P} = (\bar{p}_1, \cdots, \bar{p}_m)$, the loss function becomes $\mathcal{L}(\mathsf{P}) = \sum_i \bar{p}_i$, which is linear in P. This implies that for any P and P',

$$\mathcal{L}(\lambda\mathsf{P} + (1-\lambda)\mathsf{P}') = \lambda\mathcal{L}(\mathsf{P}) + (1-\lambda)\mathcal{L}(\mathsf{P}').$$

However, the cost $C(\mathsf{P})$ is only linear under matrix representation. We will first show that $C(\mathsf{P})$ is convex under vector representation.

Lemma 7. *For any two mechanisms $\mathsf{P} = (\bar{p}_1, \cdots, \bar{p}_m)$ and $\mathsf{P}' = (\bar{p}'_1, \cdots, \bar{p}'_m)$,*

$$C(\lambda\mathsf{P} + (1-\lambda)\mathsf{P}') \leq \lambda \cdot C(\mathsf{P}) + (1-\lambda)C(\mathsf{P}').$$

Proof. Let us denote $C(X, \mathsf{P})$ as the cost of matching some input X using mechanism P. By definition,

$$C(\mathsf{P}) = \sum_{X_i} \Pr(X_i) C(X_i, \mathsf{P}).$$

If we can show that

$$C(X, \lambda\mathsf{P} + (1-\lambda)\mathsf{P}') \leq \lambda C(X, \mathsf{P}) + (1-\lambda)C(X, \mathsf{P}'),$$

then Lemma 7 naturally follows.

Let us denote $\bar{p}_i^\lambda = \lambda\bar{p}_i + (1-\lambda)\bar{p}'_i$ and consider how the water-filling algorithm applies to $\mathsf{P}^\lambda = (\bar{p}_1^\lambda, \cdots, \bar{p}_m^\lambda)$. Consider any input X and suppose $\mathcal{F}(X) = k$. Let

$$l = \min\{l \mid \sum_{i=k}^{l} \bar{p}_i \geq 1\}, \; l' = \min\{l \mid \sum_{i=k}^{l} \bar{p}'_i \geq 1\},$$

and $l^\lambda = \min\{l \mid \sum_{i=k}^{l} \bar{p}_i^\lambda \geq 1\}$. By the water-filling lemma (Proposition 4), P would match X to output i with probability

$$p_{ki} = \begin{cases} \bar{p}_i & \text{if } k \leq i < l \\ 1 - \sum_{j=k}^{l-1} \bar{p}_j & \text{if } i = l. \end{cases}$$

Similarly, we can define p'_{ki} and p_{ki}^λ for P' and P^λ. W.l.o.g., we suppose that $l \leq l'$. By definition, we have $l \leq l^\lambda \leq l'$. For any $i \leq l$, we have $p_{ki}^\lambda = \lambda p_{ki} + (1-\lambda)p'_{ki}$. Therefore, for P^λ, $C(X, \mathsf{P}^\lambda)$ can be rewritten as

$$\sum_{i=k}^{l^\lambda} c_{ki} p_{ki}^\lambda = \sum_{i=k}^{l} c_{ki}(\lambda p_{ki} + (1-\lambda)p'_{ki}) + \sum_{i=l+1}^{l^\lambda} c_{ki} p_{ki}^\lambda$$
$$= \lambda \sum_{i=k}^{l} c_{ki} p_{ki} + (1-\lambda) \sum_{i=k}^{l} c_{ki} p'_{ki} + \sum_{i=l+1}^{l^\lambda} c_{ki} p_{ki}^\lambda.$$

The first term is actually $\lambda C(X, \mathsf{P})$. We can rewrite

$$C(X, \mathsf{P}^\lambda) - \lambda C(X, \mathsf{P}) + (1-\lambda)C(X, \mathsf{P}')$$
$$= \sum_{i=l+1}^{l'} c_{ki}(p_{ki}^\lambda - (1-\lambda)p'_{ki}).$$

Since the water-filling algorithm sets p_{ki}^λ to $\bar{p}_i^\lambda \geq (1-\lambda)\bar{p}'_i$ for all $i \in [l+1, l^\lambda)$, we have that, for any $j \in [l+1, l^\lambda]$, $\sum_{i=l+1}^{j} p_{ki}^\lambda \geq$

$(1-\lambda)\sum_{i=l+1}^{j} p'_{ki}$. Combining this with the fact that c_{ki} is increasing with i, we have

$$\sum_{i=l+1}^{l'} c_{ki}p^{\lambda}_{ki} = \sum_{i=l+1}^{l^{\lambda}} c_{ki}p^{\lambda}_{ki} \le (1-\lambda)\sum_{i=l+1}^{l^{\lambda}} c_{ki}p'_{ki}.$$

The argument here is very similar to the analyses in the proof of Theorem 1. Therefore, we have $C(X,\mathsf{P}^{\lambda}) - \lambda C(X,\mathsf{P}) + (1-\lambda)C(X,\mathsf{P}') \le 0$, which concludes our proof. □

With Lemma 6 and 7, we are ready to prove Theorem 4.

Proof. Let $C(l) = \inf\{C : (l,C) \in S\}$. By Lemma 7, $C(\cdot)$ must be a convex function. We first show that

$$C(v) = \lambda C(\lfloor v\rfloor) + (1-\lambda)C(\lceil v\rceil).$$

Suppose this is not true and $C(v) \neq \lambda C(\lfloor v\rfloor) + (1-\lambda)C(\lceil v\rceil)$. Since $C(\cdot)$ is convex, it must be that $C(v) < \lambda C(\lfloor v\rfloor)+(1-\lambda)C(\lceil v\rceil)$. This means that $(v, C(v))$ is outside the convex hull spanned by $\{(1,C(1),(2,C(2)),\cdots\}$, which also implies that $(v,C(v))$ is also not in the convex hull of S_d. We reach a contradiction here, since by Lemma 6, the convex hull of S_d is the same as the convex hull of S.

For any $\mathsf{P}_{\lfloor v\rfloor} \in \mathsf{DS}_{\lfloor v\rfloor}$ and $\mathsf{P}_{\lceil v\rceil} \in \mathsf{DS}_{\lceil v\rceil}$, by Lemma 7,

$$C(\lambda\mathsf{P}_{\lfloor v\rfloor} + (1-\lambda)\mathsf{P}_{\lceil v\rceil}) \le \lambda C(\mathsf{P}_{\lfloor v\rfloor}) + (1-\lambda)C(\mathsf{P}_{\lceil v\rceil}) = C(v).$$

This means that $\lambda\mathsf{P}_{\lfloor v\rfloor}+(1-\lambda)\mathsf{P}_{\lceil v\rceil}$ must be the optimal mechanism under budget v. This concludes our proof. □

G Further Description of Optimal Maximal Leakage Protocol

It is note that for $\log(v)$-MaxL with non-integer v, both the cost function and the privacy function are linear with respect to the elements in $\mathsf{DS}_{\lfloor v\rfloor}$ and $\mathsf{DS}_{\lceil v\rceil}$, and the optimal solution is in a weighted average of two arbitrary deterministic solutions from $\mathsf{DS}_{\lfloor v\rfloor}$ and $\mathsf{DS}_{\lceil v\rceil}$, respectively. Theorem 4 shows that it suffices to find optimal deterministic schemes in DS_v (or $\mathsf{DS}_{\lfloor v\rfloor}$ and $\mathsf{DS}_{\lceil v\rceil}$), formally described as the main protocol of Algorithm 2. In sub-algorithm 1 and 2 of Algorithm 2, we show how to find an optimal deterministic scheme by dynamic programming in $O(n\cdot m^2)$ time.

To be specific, we introduce a sub-algorithm $\mathcal{T}(i,k)$ (sub-algorithm 2 in Algorithm 2) that considers the optimal deterministic mechanism when (1) $\bar{p}_1,\cdots,\bar{p}_{i-1}$ are given and $\bar{p}_{i-1}=1$; and (2) $\sum_{j=i}^{m}\bar{p}_j$ equals k. This condition means that given the current selection of $\bar{p}_1,\cdots,\bar{p}_{i-1}$, we still need to pick k additional states within $[a:m]$. Note that there are totally $n\cdot m$ possible inputs for $\mathcal{T}(\cdot,\cdot)$. For any $\mathcal{T}(i,k)$ such that $k>0$, we consider the next state to select in the optimal scheme, i.e., the minimal $j\ge i$ such that $\bar{p}_j=1$. Once j is given, $\mathcal{T}(i,k)$ is reduced to the sub-problem $\mathcal{T}(j,k-1)$. This means that $\mathcal{T}(i,k)$ can be solved once we solve $\mathcal{T}(j,k-1)$ for all $j\ge i$. Therefore, we can use dynamic programming to solve the problem and the time complexity is $O(n\cdot m^2)$.

Suppose the optimal deterministic mechanism under the above conditions is $\{p_{ij}\}$ and its vector representation is $(\bar{p}_1,\cdots,\bar{p}_m)$, the sub-algorithm $\mathcal{T}(a,k)$ returns two outputs:

- $\mathcal{T}(a,k).Cost$: the sum of cost for any X_i such that $\mathcal{F}(X_i)\ge a$, i.e., $\sum_{(i\mid \mathcal{F}(X_i)\ge a)} q_i\sum_{j=1}^{m} c_{ij}p_{ij}$. Note that we assume $\bar{p}_1,\cdots,\bar{p}_{a-1}$ are given and $\bar{p}_{a-1}=1$, so for any X_i such that $\mathcal{F}(X_i)\le a-1$, they would be assigned to states no higher than $a-1$. To optimize the cost, it suffices to consider only $\mathcal{F}(X_i)\ge a$.
- $\mathcal{T}(a,k).Next$: the next state we select, or in other words, the smallest i such that $i\ge a$ and $\bar{p}_i=1$.

Note that $\mathcal{T}(a,k).Next$ only has $(m-a+1)$ possibilities. Therefore, we can then iterate through all possible choices and use dynamic programming to find the optimal deterministic schemes, i.e.,

$$\mathcal{T}(a,k).Next = \arg\min_{a\le a'\le m} \mathcal{T}(a'+1,k-1).Cost + \sum_{a\le\mathcal{F}(X_i)\le a'} q_i c_{ia'}.$$

After we obtain $\mathcal{T}(a,k).Next$, we can then calculate $\mathcal{T}(a,k).Cost$ straightforwardly.

H Proof of Theorem 5

We will use the following result from Theorem 4.3 in [16].

Lemma 8 ([16]). *Let* P *and* $\mathcal{Q}$ *be two continuous probability distributions on an interval* I *with finite entropy with probability density function* p *and* q*, respectively. Assume* $p(z)>0$ *for* $z\in I$*. If*

$$-\int_I q(z)\log p(z)dz = \mathsf{h}(\mathcal{P}), \tag{44}$$

then $\mathsf{h}(\mathsf{Q})\le\mathsf{h}(\mathsf{P})$*, with equality if and only if* $\mathsf{P}=\mathsf{Q}$.

By Lemma 8. we first prove the following fact: for all continuous distributions D_e supported on $[0,R]$ with second moment equaling B_0, i.e., $\int_0^R z^2\cdot\mathbb{P}(e=z)dz = B_0$, the distribution with the maximal entropy must be in a form where $p(z)=e^{-(c_1\cdot z^2+c_2)}$ for $z\in[0,R]$ with some c_1 and c_2 dependent on B_0 and R. Now, substitute such constructed P into (44), we have that for any distribution Q within $[0,R]$ and with a second moment equaling B_0,

$$\begin{aligned}-\int_0^R q(z)\log p(z)dz &= \int_0^R q(z)(c_1\cdot z^2+c_2))dz\\ &= c_1B_0+c_2.\end{aligned} \tag{45}$$

On the other hand, the entropy $\mathsf{h}(\mathsf{P})$ of constructed P equals

$$\begin{aligned}\mathsf{h}(\mathsf{P}) &= \int_0^R -p(z)\log(p(z))dz = \int_0^R (c_1\cdot z^2+c_2)\cdot p(z)dz\\ &= c_1B_0+c_2.\end{aligned} \tag{46}$$

In both (45) and (46), we use the fact that P and Q are distributions supported on $[0:R]$, i.e., $\int_0^R p(z)dz = \int_0^R q(z)dz = 1$, and are of the same second moment, i.e., $\int_0^R z^2\cdot p(z)dz = \int_0^R z^2\cdot q(z)dz = B_0$. Therefore, for distributions on $[0:R]$ with a fixed second moment, the one with the maximal entropy is with probability density in a form $p(z)\propto e^{-c_1\cdot z^2}$.

With a similar reasoning, we can also prove that for any distribution supported on $[0:R]$ with a fixed mean μ_0 and a second moment B_0, the one with the maximal entropy has a probability density function in a form $p(z)=e^{-(c_1'\cdot z^2+c_2'\cdot z+c_3')}$. For any Q with mean μ_0 and a second moment B_0, we have

$$
\begin{aligned}
&-\int_0^R q(z)\log p(z)dz = \int_0^R q(z)(c_1'\cdot z^2 + c_2'\cdot z + c_3')dz \\
&= c_1'B_0 + c_2'\mu_0 + c_3' = -\int_0^R p(z)\log p(z)dz = \mathsf{h}(\mathsf{P}).
\end{aligned} \tag{47}
$$

With the above preparation, now we go back to our objective function in (30),

$$
\min_{\mathsf{D}_e,\mathbb{E}[e^2]\le B} \mathrm{obj}(\sigma_e^2,\mathsf{D}_e) = \frac{1}{2}\cdot\left(\log(2\pi(\sigma_Y^2+\sigma_e^2))+1\right) - \mathsf{h}(e).
$$

First, it is noted $\min_{\mathsf{D}_e,\mathbb{E}[e^2]\le B}\mathrm{obj}(\sigma_e^2,\mathsf{D}_e)$ is equivalent to

$$
\min_{\sigma_e^2\in[0,B],B_0\in[0:B]}\ \min_{\mathsf{D}_e,\mathbb{E}[e^2]=B_0} \mathrm{obj}(\sigma_e^2,\mathsf{D}_e). \tag{48}
$$

It is noted that once the variance σ_e^2 and second moment B_0 of the noise e are given, by (11), the mean of the noise is also determined as $\mu_e^2 = B_0 - \sigma_e^2$. Therefore, suppose the optimal solution D_e to (30) is of a variance σ_o^2 with the second moment B_o, it consequently determines the optimal mean as $\mu_o^2 = B_o - \sigma_o^2$. Then, we know given the mean B_o and μ_o, the optimal distribution to minimize (30) (with the maximal entropy conditional on B_o and μ_o) is achievable within the class of truncated Gaussian distributions.

As a final remark, we want to mention the necessary and sufficient condition that (30) is tight for the min-max problem in (29) or when the equality of $h(Y+e) \le \frac{1}{2}\cdot\left(\log(2\pi(\sigma_Y^2+\sigma_e^2))+1\right)$ is achievable.

This is equivalent to the following question: when will there exist some distribution D_Y of Y such that for the given noise distribution D_e of e, $Y+e$ can be distributed in a Gaussian distribution when Y and e are independent. Let $\mathsf{FT}_Y(w)$ and $\mathsf{FT}_e(w)$ be the Fourier transform of Y and e, respectively. Also, let $\mathsf{FT}_G(w)$ be the Fourier transform of a Gaussian distribution with the same mean and variance as those of $Y+e$. Since the distribution of $Y+e$ is the convolution of that of Y and e, we have that $\mathsf{FT}_{Y+e}(w) = \mathsf{FT}_Y(w)+\mathsf{FT}_e(w)$. If there exists Y such that $Y+e$ is a Gaussian, i.e., $\mathsf{FT}_{Y+e}(w) = \mathsf{FT}_G(w)$, then $\mathsf{FT}_Y(w) = \mathsf{FT}_G(w)/\mathsf{FT}_e(w)$. Thus, given that Fourier transform is invertible, the sufficient and necessary condition with respect to the existence of Y becomes that $\mathsf{FT}_G(w)/\mathsf{FT}_e(w)$ is a Fourier coefficient of a distribution. By Fourier inverse theorem, this is equivalent to require the inverse of $\mathsf{FT}_G(w)/\mathsf{FT}_e(w)$ to be non-negative and this is equivalent to that $\mathsf{FT}_G(w)/\mathsf{FT}_e(w)$ needs to be positive definite functions, by Bochner's theorem.