Appendix A
Proof of Lemma 1

Lemma 1 (Parameter of Positive Laplace Noise)**.** *Suppose a processing function $\mathcal{F}: \mathcal{X}^* \to \mathbb{R}$ is such that for an arbitrary adjacent dataset pair $X \sim X'$, $|\mathcal{F}(X) - \mathcal{F}(X')| \leq \mathsf{s}$, i.e., the sensitivity of $\mathcal{F}$ is bounded by s. Then, if we select $\lambda_L = \mathsf{s}/\epsilon$ and $\mu_L \geq \mathsf{s} + \frac{s}{\epsilon} \cdot \log \frac{1}{2\delta(1-e^{-\mu_L \cdot \epsilon/s})}$ and $R = 2\mu_L$, such a (μ_L, λ_L, R)-TBL perturbation can produce (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 $\mathcal{F}(X') = \mathsf{s}$. Thus, the support domain of the distribution of $\mathcal{F}(X) + \mathsf{e}$ is $[0, R]$ while that of $\mathcal{F}(X) + \mathsf{e}$ is over $[\mathsf{s}, R + \mathsf{s}]$.

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

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

and

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

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

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

which suffices to require that

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

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

Appendix B
Proof of Theorem 3

Theorem 3. *In the above setup, if the noise produces an (ϵ, δ)-DP guarantee under sensitivity equaling 1, then for any $k \in \mathbb{Z}^+$, the optimal (minimal) k-th moment is achieved under the same noise. We define a turning point*

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

and the optimal noise is of the following form:

$$p_i = \begin{cases} \delta \cdot e^{\epsilon i} & \text{if } i < \omega \\ \delta \cdot c \cdot e^{\epsilon(2\omega - i)} & \text{if } \omega \leq i \leq \omega', \end{cases} \tag{14}$$

where 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.

To prove Theorem 3, we start with two lemmas as follows.

Lemma 4. *If a distribution $(p_0, p_1, \cdots)$ satisfies (ϵ, δ)-DP under sensitivity 1 and $p_i > \delta$, then for any $j \leq \ln(p_i/\delta)/\epsilon$, $p_{i+j} \geq e^{-j\epsilon} \cdot p_i$.*

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

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

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

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

Lemma 5. *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 $\sum_{j=0}^i p_j \geq \sum_{j=1}^i p_j$ holds for all $0 \leq i \leq n$, then for any $v_0 \leq v_1 \leq \cdots \leq v_n$, we have $\sum_{i=0}^n p_i v_i \leq \sum_{i=0}^n q_i v_i$.*

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

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

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

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

Notice that $\sum_{j=0}^i p_j \geq \sum_{j=0}^i q'_j$ 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,

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

This completes our proof. □

We will now prove Theorem 3.

Proof. We first specify how the parameters ω' and c are chosen. Recall that in Theorem 3, 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 $c(\omega) \geq e^{-2\epsilon}$, 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 $c \geq e^{-2\epsilon}$ and $p_i \geq \delta$ except for $i = \omega'$.

Let $(p_0, p_1, \cdots)$ be the distribution constructed in (14) 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 (14) is optimal. For any other distribution $(p'_0, p'_1, \cdots)$ that satisfies (ϵ, δ), we claim that for all $i \ge 0$,

$$\sum_{j=0}^{i} p_j \ge \sum_{j=0}^{i} p'_j. \tag{30}$$

This means that p is a strictly "smaller" distribution than p'. For any $m > 0$, if we set $v_i = i^m$ in Lemma 5, 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

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

By definition of (ϵ, δ)-DP, we have

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

Since p'_{-1} is undefined, $0 \in R$ and it must be that $p'_0 \le \delta = p_0$. Therefore, (30) 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 (30) holds for all $1 \le i \le \omega$.

We now consider when $\omega \le i < \omega'$. Suppose (30) holds for $i = k-1$ $(\omega \le k < \omega')$, and let us consider the scenario when $i = k$. Let us assume that (30) does not hold for $i = k$, i.e.,

$$\sum_{j=0}^{k} p_j < \sum_{j=0}^{k} p'_j. \tag{31}$$

Since (30) holds for $i = k-1$, it must be that $p_k < p'_k$. By Lemma 4, 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 (31) 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. $\square$

Appendix C
Proof of Proposition 1

Proposition 1. *Suppose T mechanisms $\mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_T$ where each $\mathcal{M}_i$ satisfies $(\alpha, \epsilon^{(i)}_{\alpha,p}, \delta^{(i)}_p)$-HRDP, then the composition of $\mathcal{M}_{[1:T]}$ satisfies (ϵ, δ)-DP, where, for any $\delta' > 0$,*

$$\epsilon \ge \sum_{i=1}^{T} \epsilon^{(i)}_{\alpha,p} + \frac{\log(1/\delta')}{\alpha - 1}, \text{ and } \delta \ge \sum_{i=1}^{T} \delta^{(i)}_p + \delta'. \tag{20}$$

We will use the following lemma mildly generalized from Proposition 10 in [37].

Lemma 6. *Let $\alpha > 1$, P and Q be two distributions defined over $\mathbb{R}$, with probability density function p and q, respectively. Let $S_d(\mathsf{Q}) = \{z | q(z) = 0\}$ be the degenerate set of Q. Then, for any $A \subset \mathbb{R}/S_d(\mathsf{Q})$,*

$$\mathsf{P}(A) \le \left(e^{\mathsf{D}_\alpha(\mathsf{P} \cdot \mathbf{1}_A \| \mathsf{Q} \cdot \mathbf{1}_A)} \cdot \mathsf{Q}(A)\right)^{(\alpha-1)/\alpha}. \tag{32}$$

Proof. Based on the Holder Inequality,

$$\begin{aligned} \mathsf{P}(A) &= \int_{z \in A} p(z) dz \\ &\le \Big(\int_{z \in A} p(z)^{\alpha} q(z)^{1-\alpha} dz\Big)^{1/\alpha} \cdot \Big(\int_{z \in A} q(z) dz\Big)^{(\alpha-1)/\alpha} \\ &= \left(e^{\mathsf{D}_\alpha(\mathsf{P} \cdot \mathbf{1}_A \| \mathsf{Q} \cdot \mathbf{1}_A)} \cdot \mathsf{Q}(A)\right)^{(\alpha-1)/\alpha}. \end{aligned} \tag{33}$$

$\square$

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

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

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

$$\begin{aligned} &\Pr(\mathcal{M}^{\otimes T}(X) \in A) \\ &= \Pr(\mathcal{M}^{\otimes T}(X) \in A_{nd}) + \Pr(\mathcal{M}^{\otimes T}(X) \in A/A_{nd}) \\ &\le T\delta_0(X) \\ &\quad + \left(e^{\mathsf{D}_\alpha(\mathbb{P}^{\otimes T}_{\mathcal{M}(X)} \cdot \mathbf{1}_{A_{nd}} \| \mathbb{P}^{\otimes T}_{\mathcal{M}(X')} \mathbf{1}_{A_{nd}})} \Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd})\right)^{\frac{\alpha-1}{\alpha}} \\ &= T\delta_0(X) + \left(e^{T\epsilon_0(X)} \Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd})\right)^{(\alpha-1)/\alpha}. \end{aligned} \tag{34}$$

In the last equation in (34), 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 [37], 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

$$\begin{aligned}&\left(e^{T\epsilon_0(X)}\Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd})\right)^{1-1/\alpha}\\ &\leq \left(e^{T\epsilon_0(X)}\Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd})\right)\cdot(\delta')^{-1/(\alpha-1)}\\ &= \left(e^{T\epsilon_0(X)+\frac{\log(1/\delta')}{\alpha-1}}\right)\cdot \Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd}).\end{aligned} \tag{35}$$

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

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

Therefore, putting things together, we obtain that

$$\begin{aligned}&\Pr(\mathcal{M}^{\otimes T}(X) \in A) \leq (T\delta_0(X)+\delta')\\ &\quad + \left(e^{T\epsilon_0(X)+\frac{\log(1/\delta')}{\alpha-1}}\right)\cdot \Pr(\mathcal{M}^{\otimes T}(X') \in A_{nd}),\end{aligned} \tag{36}$$

which provides the expression of ϵ and δ in (20), respectively.

Appendix D
Proof of Lemma 2

Lemma 2 (Contiguous Support Set of Optimal Noise)**.** *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 3, let us define

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

The key observation is that

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

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

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

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. □

Appendix E
Proof of Theorem 5

Theorem 5 (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

$$\arg\min_{\mathsf{P}_{R_0}} H(R_0, \mathsf{P}_{R_0}) \tag{37}$$
$$= \arg\min_{\mathsf{P}_{R_0}} \frac{1}{\alpha-1} \max\{T \log \sum_{i=1}^{R_0} \frac{(p_i)^\alpha}{(p_{i-1})^{\alpha-1}} + \log(\frac{1}{\delta - Tp_l}), \tag{38}$$
$$T \log \sum_{i=1}^{R_0} \frac{(p_{i-1})^\alpha}{(p_i)^{\alpha-1}} + \log(\frac{1}{\delta - Tp_r})\} \tag{39}$$
$$= \arg\min_{\mathsf{P}_{R_0}} \max\{\sum_{i=1}^{R_0} \frac{(p_i)^\alpha}{(p_{i-1})^{\alpha-1}}, \sum_{i=1}^{R_0} \frac{(p_{i-1})^\alpha}{(p_i)^{\alpha-1}}\}, \tag{40}$$

by removing the constant term and using the monotone property of $\log(\cdot)$. By the joint convexity of the Hellinger integral [43], it is known that for any two pairs of positive real numbers (p_0, q_0) and (p_1, q_1), and some $\lambda \in (0,1)$,
$$(1-\lambda)p_0^\alpha q_0^{1-\alpha} + \lambda p_1^\alpha q_1^{1-\alpha} \geq p_\lambda^\alpha q_\lambda^{1-\alpha}. \tag{41}$$
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 (40) 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. □

Appendix F
Proof of Theorem 6

Theorem 6 (Optimal Mechanism for Maximal Leakage)**.** *When v is some positive integer, the optimal solution(s) to (25) 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 (25) can be expressed as the linear interpolation of $\mathsf{DS}_{\lfloor v \rfloor}$ and $\mathsf{DS}_{\lceil v \rceil}$,*
$$\lambda \cdot \mathsf{DS}_{\lfloor v \rfloor} + (1-\lambda) \cdot \mathsf{DS}_{\lceil v \rceil}$$
$$= \{\lambda \cdot \mathsf{P}_{\lfloor v \rfloor} + (1-\lambda) \cdot \mathsf{P}_{\lceil v \rceil} \mid \mathsf{P}_{\lfloor v \rfloor} \in \mathsf{DS}_{\lfloor v \rfloor}, \mathsf{P}_{\lceil v \rceil} \in \mathsf{DS}_{\lceil v \rceil}\}.$$

Before we dive into Theorem 6, we need to first recap some of the results in [39]. 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 $\mathcal{C}(\mathsf{P})$ to denote the cost, i.e.,
$$\mathcal{L}(\mathsf{P}) = \sum_{j=1}^m \max_i p_{ij}, \quad \mathcal{C}(\mathsf{P}) = \sum_{i=1}^n q_i \sum_{j=1}^m c_{ij} p_{ij}.$$
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 $\mathcal{C}(\mathsf{P})$. 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 $(\mathcal{L}(\mathsf{P}), \mathcal{C}(\mathsf{P}))$ pairs for any mechanism P, and let S_d be the set of achievable $(\mathcal{L}(\mathsf{P}), \mathcal{C}(\mathsf{P}))$ 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 $\mathcal{C}(\mathsf{P}) = \inf[C : (\mathcal{L}(\mathsf{P}), C) \in S]$. It is shown in [39] that the boundary of the convex hull formed by S and S_d are the same.

Lemma 7. *If the cost function satisfies the requirement in Section V-A, 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 $\mathcal{C}(\mathsf{P})$ is only linear under matrix representation. We will first show that $\mathcal{C}(\mathsf{P})$ is convex under vector representation.

Lemma 8. *For any two mechanisms $\mathsf{P} = (\bar{p}_1, \cdots, \bar{p}_m)$ and $\mathsf{P}' = (\bar{p}'_1, \cdots, \bar{p}'_m)$,*
$$\mathcal{C}(\lambda\mathsf{P} + (1-\lambda)\mathsf{P}') \leq \lambda \cdot \mathcal{C}(\mathsf{P}) + (1-\lambda)\mathcal{C}(\mathsf{P}').$$

Proof. Let us denote $\mathcal{C}(X, \mathsf{P})$ as the cost of matching some input X using mechanism P. By definition,
$$\mathcal{C}(\mathsf{P}) = \sum_{X_i} \Pr(X_i)\mathcal{C}(X_i, \mathsf{P}).$$
If we can show that
$$\mathcal{C}(X, \lambda\mathsf{P} + (1-\lambda)\mathsf{P}') \leq \lambda\mathcal{C}(X, \mathsf{P}) + (1-\lambda)\mathcal{C}(X, \mathsf{P}'),$$
then Lemma 8 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
$$l = \min\{l \mid \sum_{i=k}^l \bar{p}_i \geq 1\}, \quad l' = \min\{l \mid \sum_{i=k}^l \bar{p}'_i \geq 1\},$$
and $l^\lambda = \min\{l \mid \sum_{i=k}^l \bar{p}_i^\lambda \geq 1\}$. By the water-filling lemma (Lemma 3), 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 $i \le l$, we have $p^\lambda_{ki} = \lambda p_{ki} + (1-\lambda)p'_{ki}$. Therefore, for P^λ, $\mathcal{C}(X, \mathsf{P}^\lambda)$ can be rewritten as

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

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

$$\begin{aligned}&\mathcal{C}(X, \mathsf{P}^\lambda) - \lambda\mathcal{C}(X, \mathsf{P}) + (1-\lambda)\mathcal{C}(X, \mathsf{P}') \\ = &\sum_{i=l+1}^{l'} c_{ki}(p^\lambda_{ki} - (1-\lambda)p'_{ki}).\end{aligned}$$

Since the water-filling algorithm sets p^λ_{ki} to $\bar{p}^\lambda_i \ge (1-\lambda)\bar{p}'_i$ for all $i \in [l+1, l^\lambda)$, we have that, for any $j \in [l+1, l^\lambda]$, $\sum_{i=l+1}^{j} p^\lambda_{ki} \ge (1-\lambda)\sum_{i=l+1}^{j} p'_{ki}$. 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 3. Therefore, we have $\mathcal{C}(X, \mathsf{P}^\lambda) - \lambda\mathcal{C}(X, \mathsf{P}) + (1-\lambda)\mathcal{C}(X, \mathsf{P}') \le 0$, which concludes our proof. □

With Lemma 7 and 8, we are ready to prove Theorem 6.

Proof. Let $C(l) = \inf\{C : (l, C) \in S\}$. By Lemma 8, $C(\cdot)$ 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 $C(v) \ne \lambda C(\lfloor v \rfloor) + (1-\lambda)C(\lceil v \rceil)$. 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 7, 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 8,

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

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. □

Appendix G
Full Algorithm of the Optimal Maximal Leakage Protocol

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 6 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 4, 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 Appendix G) 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.,

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

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

Appendix H
Proof of Theorem 7

We will use the following result from Theorem 4.3 in [44].

Lemma 9 ([44]). *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*

$$-\int_I q(z)\log p(z)dz = \mathsf{h}(\mathcal{P}), \quad (42)$$

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

By Lemma 9, 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

Algorithm 4 Optimal Mechanism for Maximal Leakage

Input: Objective processing function $\mathcal{F} : \mathcal{X}^* = \{X_1, X_2, \cdots, X_N\} \rightarrow \{1, 2, \cdots, m\}$, 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

into (42), we have that for any distribution Q within $[0, R]$ and with a second moment equaling B_0,

$$-\int_0^R q(z)\log p(z)dz = \int_0^R q(z)(c_1 \cdot z^2 + c_2))dz \\ = c_1 B_0 + c_2. \tag{43}$$

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

$$\mathsf{h}(\mathsf{P}) = \int_0^R -p(z)\log(p(z))dz = \int_0^R (c_1 \cdot z^2 + c_2) \cdot p(z)dz \\ = c_1 B_0 + c_2. \tag{44}$$

In both (43) and (44), 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

$$-\int_0^R q(z)\log p(z)dz = \int_0^R q(z)(c_1' \cdot z^2 + c_2' \cdot z + c_3')dz \\ = c_1' B_0 + c_2' \mu_0 + c_3' = -\int_0^R p(z)\log p(z)dz = \mathsf{h}(\mathsf{P}). \tag{45}$$

With the above preparation, now we go back to our objective function in (27),

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

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

$$\min_{\sigma_e^2 \in [0,B], B_0 \in [0:B]} \min_{\mathsf{D}_e, \mathbb{E}[e^2] = B_0} \mathsf{obj}(\sigma_e^2, \mathsf{D}_e). \tag{46}$$

It is noted that once the variance σ_e^2 and second moment B_0 of the noise e is given, by (9), 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 (27) is of a variance σ_o^2 with the second moment B_o, which 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 (27) (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 (27) is tight for the min-max problem in (26) or when the equality of $h(Y + e) \leq \frac{1}{2} \cdot \left(\log(2\pi(\sigma_Y^2 + \sigma_e^2)) + 1\right)$ is achievable.

This is equivalent to the following question when there exists 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 $\mathsf{FT}_G(w)/\mathsf{FT}_e(w)$. needs 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.