Verifying Snapping Mechanism

December 13, 2019

In order to verify the differential privacy proeprty of the snapping mechanism[3], we follow the logic rules designed from [1].

Some new rules are added into this logic in Figure 1 following with correctness proof. Then we formalized the snapping mechanism and verified its differential privacy property under these logic rules.

1 Program Logic

Definition 1 (Laplce mechanism [2])
Let $\epsilon > 0$. The Laplace mechanism $\mathcal{L}_\epsilon$: $\mathbb{R} \to \mathsf{Distr}(\mathbb{R})$ is defined by $\mathcal{L}(t) = t + v$, where $v \in \mathbb{R}$ is drawn from the Laplace distri bution $\mathsf{Laplce}(\frac{1}{\epsilon})$.

Definition 2
Let $\epsilon \leq 0$. The ϵ-DP divergence $\Delta_\epsilon(\mu_1, \mu_2)$ between two sub-distributions $\mu_1 \in \mathsf{Distr}(U)$, $\mu_2 \in \mathsf{Distr}(U)$ is defined as:

$$\underset{E \in U}{sup}\Big(\Pr_{x \leftarrow \mu_1}[x \in E] - \exp(\epsilon) \Pr_{x \leftarrow \mu_2}[x \in E]]\Big)$$

Definition 3 ((ϵ, δ) - lifting [1])
Two sub-distributions $\mu_1 \in \mathsf{Distr}(U_1)$, $\mu_2 \in \mathsf{Distr}(U_2)$are related by the (ϵ, δ) - dilation lifting of $\Psi \subseteq U_1 \times U_2$, written $\mu_1 \Psi^{\#(\epsilon,\delta)} \mu_2$, if there exist two witness sub-distributions $\mu_L \in \mathsf{Distr}(U_1 \times U_2)$ and $\mu_R \in \mathsf{Distr}(U_1, U_2)$ s.t.:

1. $\pi_1(\mu_L) = \mu_1$ and $\pi_2(\mu_R) = \mu_2$;
2. $\mathsf{supp}(\mu_L) \subseteq \Psi$ and $\mathsf{supp}(\mu_R) \subseteq \Psi$; and
3. $\Delta_\epsilon(\mu_L, \mu_R) \leq \delta$.

The logic rules we are using in our work is presented in Figure 1. The correctness of rules is proved in Theorem 1 and Theorem 2

Theorem 1
Let $\mu_1 \in \mathsf{Distr}(\mathbb{R})$, $\mu_2 \in \mathsf{Distr}(\mathbb{R})$ are defined:

$$\mu_1(x) = \mathsf{unif}(x)$$

$$\mu_2(y) = \mathsf{unif}(y)$$

$$\overline{\vdash u_1 \xleftarrow{\$} \mu \sim_{\epsilon,0} u_2 \xleftarrow{\$} \mu : \top \Rightarrow e^{-\epsilon} u_2 \le u_1 \le e^{\epsilon} u_2} \text{ AxUnif}$$

$$\overline{\vdash y_1 \xleftarrow{\$} \mathcal{L}_\epsilon(e_1) \sim_{k'\cdot\epsilon,0} y_2 \xleftarrow{\$} \mathcal{L}_\epsilon(e_2) : |k + e_1 - e_2| \le k' \Rightarrow y_1 + k = y_2} \text{ LapGen}$$

$$\overline{\vdash y_1 \xleftarrow{\$} \mathcal{L}_\epsilon(e_1) \sim_{0,0} y_2 \xleftarrow{\$} \mathcal{L}_\epsilon(e_2) : \top \Rightarrow y_1 - y_2 = e_1 - e_2} \text{ LapNull}$$

$$\overline{\vdash y_1 \xleftarrow{\$} \mu \sim_{0,0} y_2 \xleftarrow{\$} \mu : \top \Rightarrow y_1 = y_2} \text{ AxNull}$$

$$\frac{p_1 \sim_{k,0} p_2 : \Phi_1 \Rightarrow \Phi_1' \qquad c_1 \sim_{k',0} c_2 : \Phi_1' \Rightarrow \Phi_2}{\vdash p_1; c_1 \sim_{k+k',0} p_2; c_2 : \Phi_1 \Rightarrow \Phi_2} \text{ Comp}$$

Figure 1: Logic Rules from [1]

where unif is uniform distribution over $[0,1)$ whoes pdf. is defined as:

$$\mathsf{pdf}_{\mathsf{unif}}(x) = \begin{cases} 1 & x \in [0,1) \\ 0 & o.w. \end{cases}.$$

Then, $\mu_1 \Psi^{\#(\epsilon,0)} \mu_2$, where

$$\Psi = \{(x,y) \in \mathbb{R} \times \mathbb{R} | x \cdot e^{-\epsilon} \le y \le x \cdot e^{\epsilon}\}$$

Theorem 2
For any distributions $\mu_1 \in \mathsf{Distr}(\mathbb{R})$, $\mu_2 \in \mathsf{Distr}(\mathbb{R})$, $\mu_1 \Psi^{\#(0,0)} \mu_2$, where

$$\Psi = \{(x,y) \in \mathbb{R} \times \mathbb{R} | x = y\}$$

Proof of Theorem 1. Existing $\mu_L, \mu_R \in \mathsf{Distr}(\mathbb{R} \times \mathbb{R})$:

$$\mu_L(x,y) = \begin{cases} \mathsf{unif}(x) & x \cdot e^{-\epsilon} = y \wedge x \in [0,1) \\ 0 & o.w. \end{cases} \quad \mu_R(x,y) = \begin{cases} \mathsf{unif}(y) & x \cdot e^{-\epsilon} = y \wedge y \in [0,1) \\ 0 & o.w. \end{cases}.$$

Their pdf. are defined:

$$\mathsf{pdf}_{\mu_L}(x,y) = \begin{cases} \mathsf{pdf}_{\mathsf{unif}}(x) & x \cdot e^{-\epsilon} = y \wedge x \in [0,1) \\ 0 & o.w. \end{cases}$$

$$\mathsf{pdf}_{\mu_R}(x,y) = \begin{cases} \mathsf{pdf}_{\mathsf{unif}}(y) & x \cdot e^{-\epsilon} = y \wedge y \in [0,1) \\ 0 & o.w. \end{cases}.$$

- $\mathsf{supp}(\mu_L) \in \Psi \wedge \mathsf{supp}(\mu_R) \in \Psi$
  - $\mathsf{supp}(\mu_L) \subseteq \Psi$
    By definition of the pdf of μ_L, we have: $\Pr_{(x,y) \xleftarrow{\$} \mu_L} [(x,y) \notin \Psi] = 0$.

    Then we can derive $\mathsf{supp}(\mu_L) \in \Psi$

- $\mathsf{supp}(\mu_R) \subseteq \Psi$
  By definition of the pdf of μ_R, we have: $\Pr_{(x,y) \xleftarrow{\$} \mu_R}[(x,y) \notin \Psi] = 0.$
  Then we can derive $\mathsf{supp}(\mu_L) \in \Psi$

- $\pi_1(\mu_L) = \mu_1 \wedge \pi_2(\mu_R) = \mu_2$

  - $\pi_1(\mu_L) = \mu_1$
    By definition of the π_1 and pdf of μ_L, we have $\forall x \in \mathbb{R}$:

$$\mathsf{pdf}_{\pi_1(\mu_L)}(x) = \begin{cases} \int_y \mathsf{pdf}_{\mathsf{unif}}(x) & (x,y) \in \Psi \wedge x \in [0,1) \\ 0 & o.w. \end{cases} = \begin{cases} \mathsf{pdf}_{\mathsf{unif}}(x) & x \in [0,1) \\ 0 & o.w. \end{cases} = \mathsf{pdf}_{\mu_1}(x)$$

  - $\pi_1(\mu_R) = \mu_2$
    Equivalent to show$\mathsf{pdf}_{\pi_2(\mu_R)} = \mathsf{pdf}_{\mu_2}$.
    By definition of the π_2 and pdf of μ_R, we have $\forall y \in \mathbb{R}$:

$$\mathsf{pdf}_{\pi_2(\mu_R)}(y) = \begin{cases} \int_x \mathsf{pdf}_{\mathsf{unif}}(y) & (x,y) \in \Psi \wedge y \in [0,1) \\ 0 & o.w. \end{cases} = \begin{cases} \mathsf{pdf}_{\mathsf{unif}}(y) & y \in [0,1) \\ 0 & o.w. \end{cases} = \mathsf{pdf}_{\mu_2}(y)$$

- $\Delta_\epsilon(\mu_L, \mu_R) \leq 0$
  By definition of $\epsilon-$DP divergence, we have:

$$\begin{aligned} \Delta_\epsilon(\mu_L, \mu_R) &= \mathsf{Sup}_S\Big(\Pr_{(x,y) \xleftarrow{\$} \mu_L}[(x,y) \in S] - e^\epsilon \Pr_{(x,y) \xleftarrow{\$} \mu_R}[(x,y) \in S]\Big) \\ &= \mathsf{Sup}_S\Big(\int_{(x,y)\in S} \mathsf{pdf}_{\mu_L}(x,y) - e^\epsilon \int_{(x,y)\in S} \mathsf{pdf}_{\mu_R}(x,y)\Big) \end{aligned}$$

case $S \subseteq \{(x,y) | x \in [0,1) \wedge x \cdot e^{-\epsilon} = y\}$:

$$\begin{aligned} \Delta_\epsilon(\mu_L, \mu_R) &= \int_{(x,y)\in S} \mathsf{pdf}_{\mathsf{unif}}(x) - e^\epsilon \int_{(x,y)\in S} \mathsf{pdf}_{\mathsf{unif}}(y) \\ &= \int_{(x,y)\in S} \mathsf{pdf}_{\mathsf{unif}}(x) - e^\epsilon \int_{(x,y)\in S} \mathsf{pdf}_{\mathsf{unif}}(x * e^{-\epsilon}) \\ &= \int_{(x,y)\in S} \mathsf{pdf}_{\mathsf{unif}}(x) - e^\epsilon * e^{-\epsilon} \int_{(x,y)\in S} \mathsf{pdf}_{\mathsf{unif}}(x) \\ &= 0 \end{aligned}$$

case $S \subseteq \{(x,y) | x \in [1, e^\epsilon) \wedge x \cdot e^{-\epsilon} = y\}$:

$$\begin{aligned} \Delta_\epsilon(\mu_L, \mu_R) &= 0 - e^\epsilon \int_{(x,y)\in S} \mathsf{pdf}_{\mathsf{unif}}(y) \\ &< 0 \end{aligned}$$

case o.w.

$$\Delta_\epsilon(\mu_L, \mu_R) \quad = 0 - 0 = 0$$

$\square$

$$\frac{}{u_1 \xleftarrow{\$} \mu \sim_{\epsilon,0} u_2 \xleftarrow{\$} \mu : \top \Rightarrow e^{-\epsilon} u_2 \leq u_1 \leq e^{\epsilon} u_2} \textsc{AxUnif}$$

$$\frac{}{y_1 = \frac{\ln(u_1)}{\epsilon} \sim_{0,0} y_2 = \frac{\ln(u_2)}{\epsilon} : e^{-\epsilon} u_2 \leq u_1 \leq e^{\epsilon} u_2 \Rightarrow y_2 - 1 \leq y_1 \leq 1 + y_2} \textsc{AxNull}$$

$$\frac{}{s_1 \xleftarrow{\$} \{-1, 1\} \sim_{0,0} s_2 \xleftarrow{\$} \{-1, 1\} : \top \Rightarrow s_1 = s_2} \textsc{AxNull}$$

$$\frac{}{z_1 = s_1 * y_1 \sim_{0,0} z_2 = s_2 * y_2 : s_1 = s_2 \wedge y_2 - 1 \leq y_1 \leq 1 + y_2 \Rightarrow |z_1 - z_2| \leq 1} \textsc{AxNull}$$

$$\frac{}{x_1 = f(a_1) \sim_{0,0} x_2 = f(a_2) : a_1 = a_2 + 1 \wedge f(a_1) = f(a_2) + 1 \Rightarrow x_1 = x_2 + 1} \textsc{AxNull}$$

$$\frac{}{w_1 = x_1 + z_1 \sim_{0,0} w_2 = x_2 + z_2 : x_1 = x_2 + 1 \wedge |z_1 - z_2| \leq 1 \wedge -2 \leq k \leq 0 \Rightarrow w_1 + k = w_2} \textsc{AxNull}$$

$$\frac{}{w'_1 = \lfloor w_1 \rceil_\Lambda \sim_{0,0} w'_2 = \lfloor w_2 \rceil_\Lambda : w_1 + k = w_2 \wedge -2 \leq k \leq 0 \Rightarrow w'_1 + k = w'_2} \textsc{AxNull}$$

$$\frac{}{r_1 = \mathsf{clamp}_B(w'_1) \sim_{0,0} r_2 = \mathsf{clamp}_B(w'_2) : w'_1 + k = w'_2 \wedge -2 \leq k \leq 0 \Rightarrow r_1 + k = r_2} \textsc{AxNull}$$

$$\frac{\cdots}{r_1 = \mathsf{Snap}(a_1) \sim_{\epsilon,0} r_2 = \mathsf{Snap}(a_1) : a_1 = a_2 + 1 \wedge f(a_1) = f(a_2) + 1 \wedge |k + f(a_1) - f(a_2)| \leq 1 \Rightarrow r_1 + k = r_2} \textsc{Comp}$$

Figure 2: Coupling Derivation of two Snap mechanisms: $\mathsf{Snap}(a_1)$, $\mathsf{Snap}(a_2)$

2 Formalization of Snap Mechanism in Programming Logic

Definition 4 ($\mathsf{Snap}(a) : A \to \mathsf{Distr}(B)$)
The ideal Snapping mechanism $\mathsf{Snap}(a)$ is defined as:

$$u \xleftarrow{\$} \mu; y = \frac{\ln(u)}{\epsilon}; s \xleftarrow{\$} \{-1, 1\}; z = s * y; x = f(a); w = x + z; w' = \lfloor w \rceil_\Lambda; r = \mathsf{clamp}_B(w')$$

where f is the query function over input $a \in A$, ϵ is the privacy budget, B is the clampping bound and Λ is the rounding argument satisfying $\lambda = 2^k$ where 2^k is the smallest power of 2 greater or equal to the $\frac{1}{\epsilon}$.

Theorem 3
The Snap mechanism is differentially praivate by following derivation in Figure 2.

Definition 5 (ϵ **- dilation)**

.

Let $\epsilon \geq 0$. The ϵ-dilation $D_\epsilon(\mu_1, \mu_2)$ between two sub-distributions $\mu_1 \in \mathsf{Distr}(U)$, $\mu_2 \in \mathsf{Distr}(U)$is defined as:

$$\underset{E \in U}{sup}\Big(\Pr_{x \leftarrow \mu_1}[x \in E] - \exp(\epsilon) \Pr_{x \leftarrow \mu_2}[x \in \exp(-\epsilon) \cdot E]]\Big)$$

Proposition 4 **((ϵ,δ)-differential privacy)**
For every pair of sub-distributions $\mu_1 \in \mathsf{Distr}(U)$, $\mu_2 \in \mathsf{Distr}(U)$, s.t.

$$D_\epsilon(\mu_1,\mu_2) \le \delta,$$

The snapping mechanism $\mathsf{Snap}(\mu,a) : \mathsf{Distr}(U) \to A \to \mathsf{Distr}(B)$ is (ϵ,δ) - differentially private w.r.t. an adjacency relation Φ for every two adjacent inputs a, a' and μ_1,μ_2

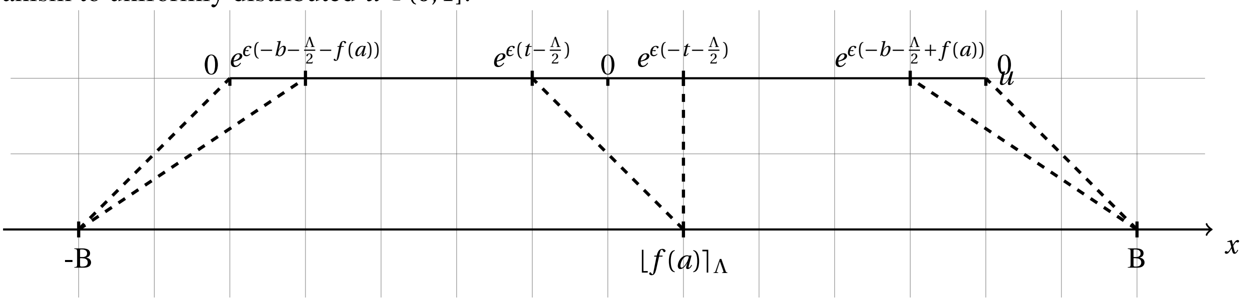
Proof. Followed directly by unfolding the Snap mechanism.

$$\begin{aligned}
\Pr_{x \leftarrow \mathsf{Snap}(\mu_1,a)}[x=e] &= \Pr_{u\leftarrow\mu_1}[\lfloor f(a) + \tfrac{S\cdot\log(u)}{\epsilon}\rceil_\Lambda = e] \\
&= \Pr_{u\leftarrow\mu_1}[u \in [\tfrac{\exp((e-\frac{\Lambda}{2}-f(a))\epsilon)}{S}, \tfrac{\exp((e+\frac{\Lambda}{2}-f(a))\epsilon)}{S})] \\
&\le \exp(\epsilon) \Pr_{u\leftarrow\mu_2}[u \in \exp(-\epsilon)[\tfrac{\exp((e-\frac{\Lambda}{2}-f(a))\epsilon)}{S}, \tfrac{\exp((e+\frac{\Lambda}{2}-f(a))\epsilon)}{S})] \\
&= \exp(\epsilon) \Pr_{u\leftarrow\mu_2}[\lfloor f(a') + \tfrac{S\cdot\log(u)}{\epsilon}\rceil_\Lambda = e] \\
&= \exp(\epsilon) \Pr_{x\leftarrow\mathsf{Snap}(\mu_2,a')}[x=e]
\end{aligned}$$

□

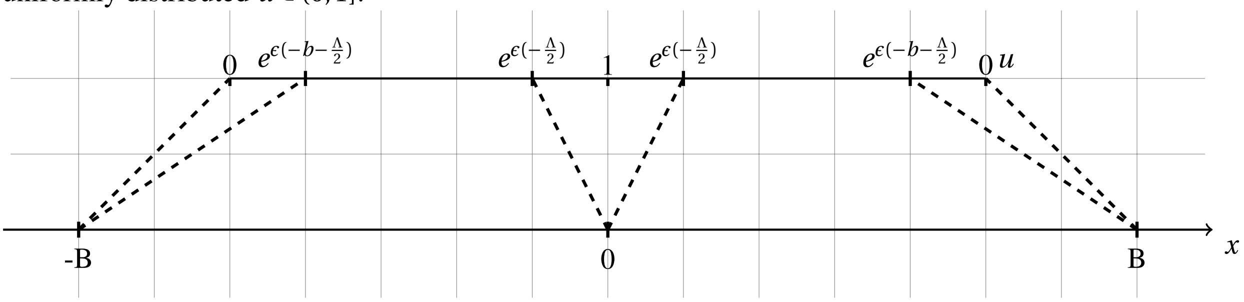
3 Proof of Differential Privacy for Snap Mechanism

Assume x be the output of Snap mechanism, we have following maps from the output of Snap mechanism to uniformly distributed $u \in (0,1]$.



where b is the greatest rounding of Λ that is smaller than B and $t = \lfloor f(a)\rceil_\Lambda - f(a)$.

Given the $f(a) = \lfloor f(a)\rceil_\Lambda = 0$, we have following maps from the output of Snap mechanism to uniformly distributed $u \in (0,1]$.



Assuming that $f(a) \in [-B,B]$, otherwise we can always redefine the $f(a)$ restricting its output in this range. The probability of obtaining output x from Snap mechanism can be calculated by cases of x:

case $x = -B$

In this case, we know $s = 1$.
We have: $\lfloor f(a) + \frac{1}{\epsilon}\ln(u) \rceil_\Lambda \leq -B$.
Since b is the greatest rounding of Λ that is smaller than B, then $-b$ is the smallest rounding of Λ that is greater than $-B$, we have: $f(a) + \frac{1}{\epsilon}\ln(u) < -b - \frac{\Lambda}{2}$.
Then we get: $u \in \left(0, e^{\epsilon(-b-\frac{\Lambda}{2}-f(a))}\right)$

case $x \in (-B, \lfloor f(a) \rceil_\Lambda)$

In this case, we know $s = 1$ and $x \in \left[-b, \lfloor f(a) \rceil_\Lambda - \Lambda\right]$.
We have: $\lfloor f(a) + \frac{1}{\epsilon}\ln(u) \rceil_\Lambda = x$.
By the rule of rounding, we get: $u \in \left[e^{\epsilon(x-\frac{\Lambda}{2}-f(a))}, e^{\epsilon(x+\frac{\Lambda}{2}-f(a))}\right)$.
By the range of x, we get: $u \in \left[e^{\epsilon(-b-\frac{\Lambda}{2}-f(a))}, e^{\epsilon(t-\frac{\Lambda}{2})}\right)$.

case $x = \lfloor f(a) \rceil_\Lambda$

subcase $s = 1$
In this case, we have: $\lfloor f(a) + \frac{1}{\epsilon}\ln(u) \rceil_\Lambda = \lfloor f(a) \rceil_\Lambda$.
Then we can get: $u \in \left[e^{\epsilon(t-\frac{\Lambda}{2})}, e^{\epsilon(t+\frac{\Lambda}{2})}\right]$.
Since $t = \lfloor f(a) \rceil_\Lambda - f(a)$, we know: $-\frac{\Lambda}{2} \leq t \leq \frac{\Lambda}{2}$. So we can get: $e^{\epsilon(t+\frac{\Lambda}{2})} > 1$.
Since $u \in (0, 1]$, we have: $u \in \left[e^{\epsilon(t-\frac{\Lambda}{2})}, 1\right]$.

subcase $s = -1$
By the symmetric of the range, we can get: $u \in \left[e^{\epsilon(-t-\frac{\Lambda}{2})}, 1\right]$.

case $x \in (\lfloor f(a) \rceil_\Lambda, B)$

In this case, we know $s = -1$ and $x \in \left[\lfloor f(a) \rceil_\Lambda + \Lambda, b\right]$.
We have: $\lfloor f(a) - \frac{1}{\epsilon}\ln(u) \rceil_\Lambda = x$.
By the rule of rounding, we get: $u \in \left(e^{\epsilon(f(a)-\frac{\Lambda}{2}-x)}, e^{\epsilon(f(a)+\frac{\Lambda}{2}-x)}\right]$.
By the range of x, we get: $u \in \left(e^{\epsilon(-b-\frac{\Lambda}{2}+f(a))}, e^{\epsilon(-t-\frac{\Lambda}{2})}\right]$.

case $x = B$

In this case, we know $s = -1$.
We have: $\lfloor f(a) - \frac{1}{\epsilon}\ln(u) \rceil_\Lambda \geq B$.
Since b is the greatest rounding of Λ that is smaller than B, we have: $f(a) - \frac{1}{\epsilon}\ln(u) \geq b + \frac{\Lambda}{2}$.
Then we get: $u \in \left(0, e^{\epsilon(-b-\frac{\Lambda}{2}+f(a))}\right)$

References

[1] Gilles Barthe, Marco Gaboardi, Benjamin Grégoire, Justin Hsu, and Pierre-Yves Strub. Proving differential privacy via probabilistic couplings. In *LICS 2016*.

[2] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating Noise to Sensitivity in Private Data Analysis. In *TCC*, 2016.

[3] Ilya Mironov. On significance of the least significant bits for differential privacy. In *CCS 2012*, 2012.