

# Verifying Snapping Mechanism

November 12, 2019

In order to verify the differential privacy property of snapping mechanism, 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 Logic Rules

### Definition 1 (Laplace mechanism [2])

Let  $\epsilon > 0$ . The Laplace mechanism  $\mathcal{L}_\epsilon: \mathbb{R} \rightarrow \text{Distr}(\mathbb{R})$  is defined by  $\mathcal{L}(t) = t + v$ , where  $v \in \mathbb{R}$  is drawn from the Laplace distribution  $\text{Laplace}(\frac{1}{\epsilon})$ .

### Definition 2

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

$$\sup_{E \in \mathcal{U}} \left( \Pr_{x \leftarrow \mu_1} [x \in E] - \exp(\epsilon) \Pr_{x \leftarrow \mu_2} [x \in E] \right)$$

### Definition 3 ( $(\epsilon, \delta)$ - lifting [1])

Two sub-distributions  $\mu_1 \in \text{Distr}(U_1)$ ,  $\mu_2 \in \text{Distr}(U_2)$  are related by the  $(\epsilon, \delta)$  - 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  $\mu_L \in \text{Distr}(U_1 \times U_2)$  and  $\mu_R \in \text{Distr}(U_1, U_2)$  s.t.:

1.  $\pi_1(\mu_L) = \mu_1$  and  $\pi_2(\mu_R) = \mu_2$ ;
2.  $\text{supp}(\mu_L) \subseteq \Psi$  and  $\text{supp}(\mu_R) \subseteq \Psi$ ; 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  $\mu_1 \in \text{Distr}(\mathbb{R})$ ,  $\mu_2 \in \text{Distr}(\mathbb{R})$  are defined:

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

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

$$\begin{array}{c}
\frac{}{\vdash u_1 \xleftarrow{\$} \mu \sim_{\epsilon,0} u_2 \xleftarrow{\$} \mu : \top \Rightarrow e^{-\epsilon} u_2 \leq u_1 \leq e^{\epsilon} u_2} \text{AXUNIF} \\
\\
\frac{}{\vdash y_1 \xleftarrow{\$} \mathcal{L}_{\epsilon}(e_1) \sim_{k',\epsilon,0} y_2 \xleftarrow{\$} \mathcal{L}_{\epsilon}(e_2) : |k + e_1 - e_2| \leq k' \Rightarrow y_1 + k = y_2} \text{LAPGEN} \\
\\
\frac{}{\vdash y_1 \xleftarrow{\$} \mathcal{L}_{\epsilon}(e_1) \sim_{0,0} y_2 \xleftarrow{\$} \mathcal{L}_{\epsilon}(e_2) : \top \Rightarrow y_1 - y_2 = e_1 - e_2} \text{LAPNULL} \\
\\
\frac{}{\vdash y_1 \xleftarrow{\$} \mu \sim_{0,0} y_2 \xleftarrow{\$} \mu : \top \Rightarrow y_1 = y_2} \text{AXNULL} \\
\\
\frac{p_1 \sim_{k,0} p_2 : \Phi_1 \Rightarrow \Phi'_1 \quad c_1 \sim_{k',0} c_2 : \Phi'_1 \Rightarrow \Phi_2}{\vdash p_1; c_1 \sim_{k+k',0} p_2; c_2 : \Phi_1 \Rightarrow \Phi_2} \text{COMP}
\end{array}$$

Figure 1: Logic Rules from [1]

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

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

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

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

## Theorem 2

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

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

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

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

Their pdf. are defined:

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

- $\text{supp}(\mu_L) \in \Psi \wedge \text{supp}(\mu_R) \in \Psi$

- $\text{supp}(\mu_L) \subseteq \Psi$

By definition of the pdf of  $\mu_L$ , we have:  $\Pr_{(x,y) \xleftarrow{\$} \mu_L} [(x, y) \notin \Psi] = 0.$

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

–  $\text{supp}(\mu_R) \subseteq \Psi$

By definition of the pdf of  $\mu_R$ , we have:  $\Pr_{(x,y) \xleftarrow{\$} \mu_R} [(x,y) \notin \Psi] = 0$ .

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

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

–  $\pi_1(\mu_L) = \mu_1$

By definition of the  $\pi_1$  and pdf of  $\mu_L$ , we have  $\forall x \in \mathbb{R}$ :

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

–  $\pi_1(\mu_R) = \mu_2$

Equivalent to show  $\text{pdf}_{\pi_2(\mu_R)} = \text{pdf}_{\mu_2}$ .

By definition of the  $\pi_2$  and pdf of  $\mu_R$ , we have  $\forall y \in \mathbb{R}$ :

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

•  $\Delta_\epsilon(\mu_L, \mu_R) \leq 0$

By definition of  $\epsilon$ -DP divergence, we have:

$$\begin{aligned} \Delta_\epsilon(\mu_L, \mu_R) &= \sup_S \left( \Pr_{(x,y) \xleftarrow{\$} \mu_L} [(x,y) \in S] - e^\epsilon \Pr_{(x,y) \xleftarrow{\$} \mu_R} [(x,y) \in S] \right) \\ &= \sup_S \left( \int_{(x,y) \in S} \text{pdf}_{\mu_L}(x,y) - e^\epsilon \int_{(x,y) \in S} \text{pdf}_{\mu_R}(x,y) \right) \end{aligned}$$

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

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

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

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

**case** o.w.

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

□

$$\begin{array}{c}
\frac{}{u_1 \xleftarrow{\$} \mu \sim_{\epsilon,0} u_2 \xleftarrow{\$} \mu : \top \Rightarrow e^{-\epsilon} u_2 \leq u_1 \leq e^{\epsilon} u_2} \text{AxUNIF} \\
\\
\frac{}{y_1 = \frac{\ln(u_1)}{\epsilon} \sim_{0,0} y_2 = \frac{\ln(u_2)}{\epsilon} : e^{-\epsilon} u_2 \leq u_1 \leq e^{\epsilon} u_2 \Rightarrow y_2 - 1 \leq y_1 \leq 1 + y_2} \text{AxNULL} \\
\\
\frac{}{s_1 \xleftarrow{\$} \{-1, 1\} \sim_{0,0} s_2 \xleftarrow{\$} \{-1, 1\} : \top \Rightarrow s_1 = s_2} \text{AxNULL} \\
\\
\frac{}{z_1 = s_1 * y_1 \sim_{0,0} z_2 = s_2 * y_2 : s_1 = s_2 \wedge y_2 - 1 \leq y_1 \leq 1 + y_2 \Rightarrow |z_1 - z_2| \leq 1} \text{AxNULL} \\
\\
\frac{}{x_1 = f(a_1) \sim_{0,0} x_2 = f(a_2) : a_1 = a_2 + 1 \wedge f(a_1) = f(a_2) + 1 \Rightarrow x_1 = x_2 + 1} \text{AxNULL} \\
\\
\frac{}{w_1 = x_1 + z_1 \sim_{0,0} w_2 = x_2 + z_2 : x_1 = x_2 + 1 \wedge |z_1 - z_2| \leq 1 \wedge -2 \leq k \leq 0 \Rightarrow w_1 + k = w_2} \text{AxNULL} \\
\\
\frac{}{w'_1 = \lfloor w_1 \rfloor_{\Lambda} \sim_{0,0} w'_2 = \lfloor w_2 \rfloor_{\Lambda} : w_1 + k = w_2 \wedge -2 \leq k \leq 0 \Rightarrow w'_1 + k = w'_2} \text{AxNULL} \\
\\
\frac{}{r_1 = \text{clamp}_B(w'_1) \sim_{0,0} r_2 = \text{clamp}_B(w'_2) : w'_1 + k = w'_2 \wedge -2 \leq k \leq 0 \Rightarrow r_1 + k = r_2} \text{AxNULL} \\
\\
\frac{\dots}{r_1 = \text{Snap}(a_1) \sim_{\epsilon,0} r_2 = \text{Snap}(a_2) : a_1 = a_2 + 1 \wedge f(a_1) = f(a_2) + 1 \wedge -2 \leq k \leq 0 \Rightarrow r_1 + k = r_2} \text{COMP}
\end{array}$$

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

## 2 Formalization

**Definition 4** ( $\text{Snap}(a) : A \rightarrow \text{Distr}(B)$ )

The ideal Snapping mechanism  $\text{Snap}(a)$  is defined as:

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

where  $f$  is the query function over input  $a \in A$ ,  $\epsilon$  is the privacy budget,  $B$  is the clamping bound and  $\Lambda$  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 private by following derivation in Figure 2.

**Definition 5** ( $\epsilon$  - dilation)

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

$$\sup_{E \in U} \left( \Pr_{x \leftarrow \mu_1} [x \in E] - \exp(\epsilon) \Pr_{x \leftarrow \mu_2} [x \in \exp(-\epsilon) \cdot E] \right)$$

**Proposition 4**  $((\epsilon, \delta)$ -differential privacy)

For every pair of sub-distributions  $\mu_1 \in \text{Distr}(U)$ ,  $\mu_2 \in \text{Distr}(U)$ , s.t.

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

The snapping mechanism  $\text{Snap}(\mu, a) : \text{Distr}(U) \rightarrow A \rightarrow \text{Distr}(B)$  is  $(\epsilon, \delta)$  - differentially private w.r.t. an adjacency relation  $\Phi$  for every two adjacent inputs  $a, a'$  and  $\mu_1, \mu_2$

*Proof.* Followed directly by unfolding the Snap mechanism.

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

□

## 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.