• **SUPPORTING INFORMATION** •

Implicit privacy preservation: a framework based on data generation

Qing YANG[1,2], Cheng WANG[1,2], Teng HU[1,2], Xue CHEN[1,2] & Changjun JIANG[1,2*]

[1]*The Key Laboratory of Embedded System and Service Computing (Tongji University), Ministry of Education, Shanghai* 201804, *China;*
[2]*The National (Province-Ministry Joint) Collaborative Innovation Center for Financial Network Security, Tongji University, Shanghai* 201804, *China.*

Citation Yang Q, Wang C, Hu T, et al. Implicit privacy preservation: a framework based on data generation. Security and Safety2022; **1**:2022008. https://doi.org/10.1051/sands/2022008

Appendix

Proposition 1. Given a fixed encoder E and a fixed generator G, the optimal discriminators D_1^* and D_2^* are

$$D_1^*(\boldsymbol{x}_p|s) = \frac{P_r(\tilde{\boldsymbol{x}}_p|s)}{P_r(\tilde{\boldsymbol{x}}_p|s) + P_g(\hat{\boldsymbol{x}}_p|s)}, \tag{A1}$$

$$D_2^*(\boldsymbol{x}_p) = \frac{P_g(\hat{\boldsymbol{x}}_p|s=1)}{P_g(\hat{\boldsymbol{x}}_p|s=0) + P_g(\hat{\boldsymbol{x}}_p|s=1)}. \tag{A2}$$

Proof. Given a generator G and an encoder E, we can obtain the optimal discriminators D_1^* and D_2^* by maximizing $V_{CGAN}(G, D_1, D_2)$:

$$\begin{aligned}
&V_{CGAN}(G, D_1, D_2) \\
&= \int_{\boldsymbol{x}_p} P_r(\tilde{\boldsymbol{x}}_p|s) \log D_1(\tilde{\boldsymbol{x}}_p|s) d\boldsymbol{x}_p + \lambda \int_{\boldsymbol{x}_p} P_g(\hat{\boldsymbol{x}}_p|s=0) \log(1 - D_2(\hat{\boldsymbol{x}}_p)) d\boldsymbol{x}_p \\
&+ \lambda \int_{\boldsymbol{x}_p} P_g(\hat{\boldsymbol{x}}_p|s=1) \log D_2(\hat{\boldsymbol{x}}_p) d\boldsymbol{x}_p + \int_{\boldsymbol{x}_p} P_g(\hat{\boldsymbol{x}}_p|s) \log(1 - D_1(\hat{\boldsymbol{x}}_p|s)) d\boldsymbol{x}_p \\
&= \int_{\boldsymbol{x}_p} \Big[P_r(\tilde{\boldsymbol{x}}_p) \log D_1(\tilde{\boldsymbol{x}}_p|s) + P_g(\hat{\boldsymbol{x}}_p) \log(1 - D_1(\hat{\boldsymbol{x}}_p|s)) \Big] d\boldsymbol{x}_p \\
&+ \lambda \int_{\boldsymbol{x}_p} \Big[P_g(\hat{\boldsymbol{x}}_p|s=1) \log D_2(\hat{\boldsymbol{x}}_p) + P_g(\hat{\boldsymbol{x}}_p|s=0) \log(1 - D_2(\hat{\boldsymbol{x}}_p)) \Big] d\boldsymbol{x}_p.
\end{aligned} \tag{A3}$$

Solving the maximum value of $V_{CGAN}(G, D_1, D_2)$ can be converted into the maximum value of the following integrand function.

$$P_r(\tilde{\boldsymbol{x}}_p) \log D_1(\tilde{\boldsymbol{x}}_p|s) + P_g(\hat{\boldsymbol{x}}_p) \log(1 - D_1(\hat{\boldsymbol{x}}_p|s)). \tag{A4}$$

$$P_g(\hat{\boldsymbol{x}}_p|s=1) \log D_2(\hat{\boldsymbol{x}}_p) + P_g(\hat{\boldsymbol{x}}_p|s=0) \log(1 - D_2(\hat{\boldsymbol{x}}_p)). \tag{A5}$$

Take solving the optimal discriminator D_1 as an example. If $y = D_1(*|s)$, then Eq (A4) can be written as:

$$f(y) = a \log y + b \log(1 - y). \tag{A6}$$

When $(a, b) \in \mathbb{R}^2 \backslash \{0, 0\}$, we can use the first order to solve:

$$f^{'}(y) = 0 \Rightarrow y = \frac{a}{a + b}. \tag{A7}$$

This concludes the proof.

* Corresponding author (email: cjjiang@tongji.edu.cn)

Theorem 1. Given a fixed encoder E, the optimal discriminators D_1^* and D_2^*, there exists a global minimum for the function $C(G)$.

Proof. According to the definition of Jensen-Shannon divergence, $C(G)$ can be changed as:

$$
\begin{aligned}
C(G) &= \max_{D_1, D_2} V_{CGAN}(G, D_1, D_2) + V_{VAE}(E, G) \\
&= \mathbb{E}_{\boldsymbol{x}_p \sim P_r(\tilde{\boldsymbol{x}}_p|s)} \Big[\log D_1^*(\boldsymbol{x}_p|s) \Big] + \mathbb{E}_{\hat{\boldsymbol{x}}_p \sim P_g(\hat{\boldsymbol{x}}_p|s)} \Big[\log(1 - D_1^*(\hat{\boldsymbol{x}}_p|s)) \Big] \\
&+ \lambda \mathbb{E}_{\hat{\boldsymbol{x}}_p \sim P_g(\hat{\boldsymbol{x}}_p|s=1)} \Big[\log D_2^*(\hat{\boldsymbol{x}}_p) \Big] + \lambda \mathbb{E}_{\hat{\boldsymbol{x}}_p \sim P_g(\hat{\boldsymbol{x}}_p|s=0)} \Big[\log(1 - D_2^*(\hat{\boldsymbol{x}}_p)) \Big] + V_{VAE}(E, G) \\
&= \mathbb{E}_{\tilde{\boldsymbol{x}}_p \sim P_r(\tilde{\boldsymbol{x}}_p|s)} \left[\log \frac{P_r(\tilde{\boldsymbol{x}}_p|s)}{P_r(\tilde{\boldsymbol{x}}_p|s) + P_g(\hat{\boldsymbol{x}}_p|s)} \right] + \mathbb{E}_{\hat{\boldsymbol{x}}_p \sim P_g(\hat{\boldsymbol{x}}_p|s)} \left[\log(1 - \frac{P_r(\tilde{\boldsymbol{x}}_p|s)}{P_r(\tilde{\boldsymbol{x}}_p|s) + P_g(\hat{\boldsymbol{x}}_p|s)}) \right] \\
&+ \lambda \mathbb{E}_{\hat{\boldsymbol{x}}_p \sim P_g(\hat{\boldsymbol{x}}_p|s=0)} \left[\log(1 - \frac{P_g(\hat{\boldsymbol{x}}_p|s=1)}{P_g(\hat{\boldsymbol{x}}_p|s=0) + P_g(\hat{\boldsymbol{x}}_p|s=1)}) \right] \\
&+ \lambda \mathbb{E}_{\hat{\boldsymbol{x}}_p \sim P_g(\hat{\boldsymbol{x}}_p|s=1)} \left[\log \frac{P_g(\hat{\boldsymbol{x}}_p|s=1)}{P_g(\hat{\boldsymbol{x}}_p|s=0) + P_g(\hat{\boldsymbol{x}}_p|s=1)} \right] + V_{VAE}(E, G) \\
&= -(2+\lambda) \log 4 + 2 \times JS(P_r(\tilde{\boldsymbol{x}}_p|s) || P_g(\hat{\boldsymbol{x}}_p|s)) \\
&+ 2\lambda \times JS(P_g(\hat{\boldsymbol{x}}_p|s=0) || P_g(\hat{\boldsymbol{x}}_p|s=1)) + V_{VAE}(E, G),
\end{aligned} \tag{A8}
$$

where JS(*) represents Jensen-Shannon divergence. Since Jensen-Shannon divergence, Kullback-Leibler divergence and cross entropy are convex functions, $C(G)$ can converge to a global minimum. This concludes the proof.