Optimal Utility-Privacy Trade-off with Total Variation Distance as a Privacy Measure

Borzoo Rassouli* and Deniz Gündüz†
*School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK.
†Department of Electrical and Electronic Engineering, Imperial College London, London, UK.
Emails: b.rassouli@essex.ac.uk and d.gunduz@imperial.ac.uk

Abstract—**The total variation distance is proposed as a privacy measure in an information disclosure scenario when the goal is to reveal some information about available data in order to receive utility, while preserving the privacy of sensitive data from the legitimate receiver. The total variation distance is motivated as a measure of privacy-leakage by showing that: i) it satisfies the *post-processing* and *linkage* inequalities, which makes it consistent with an intuitive notion of a privacy measure; ii) the optimal utility-privacy trade-off can be solved through a standard linear program when total variation distance is employed as the privacy measure; iii) it provides a bound on the privacy-leakage measured by mutual information, *maximal leakage*, or the improvement in an inference attack with an arbitrary bounded cost function[1].**

Index Terms—**Privacy, total variation distance, utility-privacy trade-off**

I. Introduction

We measure, store, and share an immense amount of data about ourselves, from our vital signals to our energy consumption profile. We often disclose these data in return of various services, e.g., better health monitoring, more reliable energy grid, etc. However, with the growing power of machine learning algorithms, the data we share can reveal more accurate and detailed personal information, beyond what we are willing to share. One solution to this problem is to develop privacy-preserving data release mechanisms that can provide a trade-off between the utility we receive and the information we leak. Denoting the data to be released by random variable Y, and the latent private variable as X, we apply a *privacy-preserving mapping* on Y, whereby a distorted version of Y, denoted by U, is shared instead of Y. Typically, privacy and utility are competing goals: The more distorted version of Y is revealed, the less information can be inferred about X, while the less utility can be obtained.

The information-theoretic view of privacy has gained increasing attention recently. In [1], a general statistical inference framework is proposed to capture the loss of privacy in legitimate transactions of data. In [2], the privacy-utility trade-off under the log-loss cost function is considered, called the *privacy funnel*. In a more recent work [3], a generic privacy model is considered, where the privacy mapping has access to a noisy observation W of the pair (X,Y). Different well-known privacy measures and their characteristics are also investigated in [3].

We study the information-theoretic privacy in this paper. In particular, we measure the privacy-leakage (about the private variable X by revealing U) by the total variation distance as

$$T(X;U) \triangleq \max_{u\in\mathcal{U}} \|p_X(\cdot|u) - p_X(\cdot)\|_{TV} = \frac{1}{2}\max_{u\in\mathcal{U}} \|\mathbf{p}_{X|u} - \mathbf{p}_X\|_1, \tag{1}$$

where $\mathbf{p}_{X|u}$ and $\mathbf{p}_X$ are the probability vectors corresponding to probability mass functions (pmf) $p_{X|U}(\cdot|u)$ and $p_X(\cdot)$, respectively. First, we characterize the optimal utility-privacy trade-off under this privacy measure for three different utility measures, namely mutual information, minimum mean-square error (MMSE), and probability of error. Then, we motivate the proposed privacy measure by showing that it satisfies both the *post-processing* and *linkage* inequalities [3], and it provides a bound on the leakage measured by mutual information, *maximal leakage* [4], or the improvement in an inference attack with an arbitrary bounded cost function[2].

Notations. Random variables are denoted by capital letters, their realizations by lower case letters. Matrices and vectors are denoted by bold capital and bold lower case letters, respectively. For integers $m \leq n$, we have the discrete interval $[m:n] \triangleq \{m, m+1, \ldots, n\}$. For an integer $n \geq 1$, $\mathbf{1}_n$ denotes an n-dimensional all-one column vector. For a random variable $X \in \mathcal{X}$, with finite $|\mathcal{X}|$, the probability simplex $\mathcal{P}(\mathcal{X})$ is the standard $(|\mathcal{X}|-1)$-simplex given by

$$\mathcal{P}(\mathcal{X}) = \left\{\mathbf{v}\in\mathbb{R}^{|\mathcal{X}|} \middle| \mathbf{1}_{|\mathcal{X}|}^T \cdot \mathbf{v} = 1,\ v_i \geq 0,\ \forall i \in [1:|\mathcal{X}|]\right\}.$$

Furthermore, to each pmf on X, denoted by $p_X(\cdot)$, corresponds a probability vector $\mathbf{p}_X \in \mathcal{P}(\mathcal{X})$, whose i-th element is $p_X(x_i)$ ($i \in [1:|\mathcal{X}|]$). Likewise, for a pair of random variables (X,Y) with joint pmf $p_{X,Y}$, the probability vector $\mathbf{p}_{X|y}$ corresponds to the conditional pmf $p_{X|Y}(\cdot|y), \forall y \in \mathcal{Y}$, and $\mathbf{P}_{X|Y}$ is an $|\mathcal{X}|\times|\mathcal{Y}|$ matrix with columns $\mathbf{p}_{X|y}, \forall y\in\mathcal{Y}$. $F_Y(\cdot)$ denotes the cumulative distribution function (CDF) of random variable Y. For $0 \leq t \leq 1$, $H_b(t) \triangleq -t\log_2 t - (1-t)\log_2(1-t)$ denotes the binary entropy function with the convention $0\log 0 = 0$. Throughout the paper, for a random variable Y with the corresponding probability vector $\mathbf{p}_Y$, the

[1] This work has been carried out when the first author was with the Information Processing and Communications Laboratory at Imperial College London. This work has received funding from the European Research Council (ERC) through Starting Grant BEACON (agreement No. 677854), and from the UK Engineering and Physical Sciences Research Council (EPSRC) through the project COPES (173605884).

[2] Throughout this paper, we refer to the longer version of it provided in [5].

entropies $H(Y)$ and $H(\mathbf{p}_Y)$ are written interchangeably. For $\mathbf{x} \in \mathbb{R}^n$ and $p \in [1,\infty]$, the L^p-norm is defined as $\|\mathbf{x}\|_p \triangleq (\sum_i |x_i|^p)^{\frac{1}{p}}, p \in (1,\infty)$, and $\|\mathbf{x}\|_\infty \triangleq \max_{i\in[1:n]} |x_i|$.

II. System model and preliminaries

Consider a pair of random variables $(X,Y) \in \mathcal{X} \times \mathcal{Y}$ ($|\mathcal{X}|, |\mathcal{Y}| < \infty$) distributed according to the joint distribution $p_{X,Y}$. We assume that $p_Y(y) > 0, \forall y \in \mathcal{Y}$, and $p_X(x) > 0, \forall x \in \mathcal{X}$, since otherwise the supports $\mathcal{Y}$ or/and $\mathcal{X}$ could have been modified accordingly. Let Y denote the available data to be released, while X denote the latent private data. Assume that the *privacy mapping/data release mechanism* takes Y as input and maps it to the *released data* denoted by U. In this scenario, $X - Y - U$ form a Markov chain, and the privacy mapping is denoted by the conditional distribution $p_{U|Y}$. Let $J(X;U) \in [0,+\infty)$ be a generic privacy measure as a functional of the joint distribution $p_{X,U}$ that captures the amount of (information) leakage from X to U. Hence, the smaller $J(X;U)$ is, the higher privacy is achieved by the mapping $p_{U|Y}$. Also, let $R(Y;U) \in [0,+\infty)$ be a functional of the joint distribution $p_{Y,U}$, and denote an application-specific quantity that measures the amount of utility obtained by disclosing U. Therefore, the utility-privacy trade-off can be written as

$$\sup_{p_{U|Y}: X-Y-U,\ J(X;U)\leq\epsilon} R(Y;U). \tag{2}$$

Minimum privacy-leakage is assured when X and U are statistically independent. This happens if and only if $p_{X|U}(\cdot|u) = p_X(\cdot), \forall u \in \mathcal{U}$, or equivalently, the statistical distance between any $p_{X|U}(\cdot|u) (u \in \mathcal{U})$ and $p_X(\cdot)$ is zero. Intuitively, this motivates us to measure the privacy of a mapping $p_{U|Y}$ by the maximum statistical distance between $p_{X|U}(\cdot|u)$ and $p_X(\cdot)$, where the maximum is over $u \in \mathcal{U}$. In this paper, we use the total variation distance between $p_{X|U}(\cdot|u)$ and $p_X(\cdot)$ to measure the privacy-leakage as in (1)[3], i.e., $J(\cdot;\cdot) = T(\cdot;\cdot)$.

III. The optimal utility-privacy trade-off

In this section, we consider three utility measures and address the optimal utility-privacy trade-off problem when the privacy measure is given in (1).

A. Mutual information

Let $m_\epsilon(X,Y)$ be defined[4] as

$$m_\epsilon(X,Y) \triangleq \sup_{p_{U|Y}: X-Y-U,\ T(X;U)\leq\epsilon} I(Y;U). \tag{3}$$

Proposition 1. In the evaluation of (3), it is sufficient to have $|\mathcal{U}| \leq |\mathcal{Y}|$. Also, the supremum is achieved in (3).

Proof. The proof is provided in Appendix A. □

In the sequel, the optimal utility-privacy trade-off in (3) is characterized. To this end, we start with the special case of binary Y that admits a closed-form solution.

Theorem 1. Let $(X,Y) \in \mathcal{X} \times \{y_1, y_2\}$ ($|\mathcal{X}| < \infty$) with $p_Y(y_1) = p$ and $\mathbf{P}_{X|Y} = \begin{bmatrix}\mathbf{p}_{X|y_1} & \mathbf{p}_{X|y_2}\end{bmatrix}$. We have

$$m_\epsilon(X,Y) = H_b(p) - \frac{p-m}{M-m}H_b(M) - \frac{M-p}{M-m}H_b(m), \tag{4}$$

where

$$\begin{aligned} m &\triangleq \max\left\{0, p - \frac{2\epsilon}{\|\mathbf{p}_{X|y_1} - \mathbf{p}_{X|y_2}\|_1}\right\} \\ M &\triangleq \min\left\{1, p + \frac{2\epsilon}{\|\mathbf{p}_{X|y_1} - \mathbf{p}_{X|y_2}\|_1}\right\}. \end{aligned} \tag{5}$$

Proof. Let $p_{Y|U}(y_1|u)$ be denoted by $q_u, \forall u \in \mathcal{U}$. From the constraint $T(X;U) \leq \epsilon$, we have

$$\left\| \mathbf{P}_{X|Y} \left(\begin{bmatrix} q_u \\ 1-q_u \end{bmatrix} - \begin{bmatrix} p \\ 1-p \end{bmatrix} \right) \right\|_1 \leq 2\epsilon, \forall u \in \mathcal{U},$$

which results in

$$|q_u - p| \leq \frac{2\epsilon}{\|\mathbf{p}_{X|y_1} - \mathbf{p}_{X|y_2}\|_1} \Rightarrow q_u \in [m, M], \forall u \in \mathcal{U}.$$

Hence, $m_\epsilon(X,Y)$ is given by

$$m_\epsilon(X,Y) = H_b(p) - \min_{\substack{p_U(\cdot), q_u: \\ m\leq q_u \leq M,\ \sum_u p_U(u)q_u = p}} \sum_u p_U(u) H_b(q_u). \tag{6}$$

From the concavity of $H_b(\cdot)$, we have

$$H_b(q_u) \geq \frac{q_u - m}{M-m}H_b(M) + \frac{M-q_u}{M-m}H_b(m), \forall q_u \in [m, M]. \tag{7}$$

From (7) and (6), (4) is obtained with $\mathcal{U} = \{u_1, u_2\}, p_U(u_1) = \frac{p-m}{M-m}$, $q_{u_1} = M$ and $q_{u_2} = m$. □

Theorem 2. For a pair of random variables $(X,Y) \in \mathcal{X} \times \mathcal{Y}$ ($|\mathcal{X}|, |\mathcal{Y}| < \infty$), $m_\epsilon(X,Y)$ is the solution to a standard linear program (LP) (given in (12)).

Proof. Let $\tilde{\mathbb{S}}_\epsilon$ be defined as

$$\tilde{\mathbb{S}}_\epsilon \triangleq \left\{ \mathbf{x} \in \mathcal{P}(\mathcal{X}) \middle| \frac{1}{2}\|\mathbf{x} - \mathbf{p}_X\|_1 \leq \epsilon \right\}, \tag{8}$$

and $\mathbb{S}_\epsilon$ be the inverse image of $\tilde{\mathbb{S}}_\epsilon$ under the linear transformation $\mathbf{P}_{X|Y}$, i.e.,

$$\mathbb{S}_\epsilon \triangleq \left\{ \mathbf{x} \in \mathcal{P}(\mathcal{Y}) \middle| \frac{1}{2}\|\mathbf{P}_{X|Y}(\mathbf{x} - \mathbf{p}_Y)\|_1 \leq \epsilon \right\}. \tag{9}$$

It can be verified that $\tilde{\mathbb{S}}_\epsilon$ is a convex polytope in $\mathcal{P}(\mathcal{X})$, since it can be written as the intersection of a finite number of closed half-spaces (in $\mathcal{P}(\mathcal{X})$) of the form $\sum_{i=1}^{|\mathcal{X}|} \alpha_i x_i \leq b_\epsilon$, where $\alpha_i \in \{-1, 0, 1\}$. The same also holds for its inverse image under the linear transformation $\mathbf{P}_{X|Y}$, i.e., $\mathbb{S}_\epsilon$ is a convex polytope in $\mathcal{P}(\mathcal{Y})$ written as the intersection of a finite number of closed half-spaces (in $\mathcal{P}(\mathcal{Y})$) of the form $\sum_{i=1}^{|\mathcal{Y}|} \beta_i x_i \leq c_\epsilon$, where $\beta_i \in [-1, 1]$. Also, note the fact that $\mathbb{S}_\epsilon$ $(\in \mathcal{P}(\mathcal{Y}))$ is a bounded set with a finite number of extreme points. An example of these regions is provided in [5].

[3] Note that T is not symmetric, and we have $T(X;U) \leq T(X;X) = 1 - \min_x p_X(x)$; Also, $T(X;U) = 0$ iff X and U are independent.

[4] ϵ is assumed to be in its feasible range of $[0, 1 - \min_x p_X(x)]$. This range can be further tightened to $[0, T(X;Y)]$ from the *post-processing* inequality in Section IV.

For a mapping $p_{U|Y}$ that satisfies the Markov chain $X - Y - U$ and the constraint $T(X;U) \leq \epsilon$, we must have $\mathbf{p}_{Y|u} \in \mathbb{S}_\epsilon, \forall u \in \mathcal{U}$. On the other hand, for any mapping $p_{U|Y}$, for which $\mathbf{p}_{Y|u} \in \mathbb{S}_\epsilon,\ \forall u \in \mathcal{U}$, we can build the Markov chain $X - Y - U$, where $T(X;U) \leq \epsilon$. Therefore, the following equivalence holds for mappings $p_{U|Y}$:

$$X - Y - U,\ T(X;U) \leq \epsilon \Longleftrightarrow \mathbf{p}_{Y|u} \in \mathbb{S}_\epsilon,\ \forall u \in \mathcal{U}. \tag{10}$$

This leads us to

$$\begin{aligned} m_\epsilon(X,Y) &= \max_{\substack{p_{U|Y}: \\ \mathbf{p}_{Y|u}\in\mathbb{S}_\epsilon,\ \forall u\in\mathcal{U}}} I(Y;U) \\ &= H(Y) - \min_{\substack{p_U(\cdot), \mathbf{p}_{Y|u}\in\mathbb{S}_\epsilon,\ \forall u\in\mathcal{U}: \\ \sum_u p_U(u)\mathbf{p}_{Y|u}=\mathbf{p}_Y}} H(Y|U), \end{aligned} \tag{11}$$

where in (11), since the minimization is over $p_U(\cdot)$ and $\mathbf{p}_{Y|u}$ rather than $p_{U|Y}$, a constraint was added to preserve the marginal distribution $\mathbf{p}_Y$, which is already specified by $p_{X,Y}$.

Proposition 2. In minimizing $H(Y|U)$ over $\mathbf{p}_{Y|u} \in \mathbb{S}_\epsilon$, it is sufficient to consider only $|\mathcal{Y}|$ extreme points of $\mathbb{S}_\epsilon$.

Proof. Assume that the minimum in (11) is achieved by $N(\leq |\mathcal{Y}|)$ points in $\mathbb{S}_\epsilon$, which follows from Proposition 1. We prove that all of these N points must belong to the extreme points of $\mathbb{S}_\epsilon$. Let $\mathbf{p}$ be an arbitrary point among these N points. $\mathbf{p}$ can be written as[5] $\mathbf{p} = \sum_{i=1}^{|\mathcal{Y}|} \alpha_i \mathbf{p}_i$, where $\alpha_i \geq 0$ ($\forall i \in [1 : |\mathcal{Y}|]$) and $\sum_{i=1}^{|\mathcal{Y}|} \alpha_i = 1$; points $\mathbf{p}_i$ ($\forall i \in [1 : |\mathcal{Y}|]$) belong to the extreme points of $\mathbb{S}_\epsilon$ and $\mathbf{p}_i \neq \mathbf{p}_j$ ($i \neq j$). From the concavity of entropy, we have $H(\mathbf{p}) \geq \sum_{i=1}^{|\mathcal{Y}|} \alpha_i H(\mathbf{p}_i)$ where the equality holds if and only if all of the α_is but one are zero. From the definition of an extreme point, if $\mathbf{p}$ is not an extreme point of $\mathbb{S}_\epsilon$, it can be written with at least two non-zero α_is, which makes this inequality strict. However, this violates the assumption that the N points achieve the minimum. Hence, all of the N points of the minimizer must belong to the set of extreme points of $\mathbb{S}_\epsilon$. □

Proposition 2, which is the generalization of (7), divides the problem in (11) into two steps: in step one, the extreme points of $\mathbb{S}_\epsilon$ are identified, while in step two, proper weights over these extreme points are obtained to minimize the objective function, $H(Y|U)$.

For the first step, we proceed as follows. We can write $\mathbb{S}_\epsilon$ as the union of the sets[6] that have the general form of $\tilde{\mathbb{D}} = \{\mathbf{x} \in \mathbb{R}^{|\mathcal{Y}|} | \tilde{\mathbf{A}}\mathbf{x} \leq \mathbf{b}, \mathbf{1}_{|\mathcal{Y}|}^T \cdot \mathbf{x} = 1, \mathbf{x} \geq 0\}$. Each of these sets is a convex polytope in $\mathcal{P}(\mathcal{Y})$ whose extreme points are the basic feasible solutions (see [6], [7]) of their corresponding set $\mathbb{D} = \{\mathbf{x} \in \mathbb{R}^{|\mathcal{Y}|'} | \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\}$, where $|\mathcal{Y}| + 2 \leq |\mathcal{Y}|' \leq 2|\mathcal{Y}| + 1$. The procedure of finding the basic feasible solutions is a classical problem, which is omitted here due to lack of space. Further details and examples are provided in [5].

For the second step, we proceed as follows. Assume that the extreme points of $\mathbb{S}_\epsilon$, found in the previous step, are denoted by $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_K$. Then (11) is equivalent to

$$\begin{aligned} m_\epsilon(X,Y) = H(Y) - \min_{\mathbf{w}\geq 0} &\begin{bmatrix} H(\mathbf{p}_1) & H(\mathbf{p}_2) & \ldots & H(\mathbf{p}_K) \end{bmatrix} \cdot \mathbf{w} \\ \text{s.t. } &\begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \ldots & \mathbf{p}_K \end{bmatrix} \mathbf{w} = \mathbf{p}_Y, \end{aligned} \tag{12}$$

where $\mathbf{w}$ is a K-dimensional weight vector, and it can be verified that the constraint $\sum_{i=1}^K w_i = 1$ is satisfied if the constraint in (12) is met. The problem in (12) is a standard linear program (LP). □

B. Minimum mean-square error (MMSE)

Assume that the utility is measured by the decrease in the mean-square error, i.e., the trade-off is given by

$$\min_{\substack{p_{U|Y}: \\ X-Y-U \\ T(X;U)\leq\epsilon}} \mathbb{E}[(Y-U)^2], \tag{13}$$

where the expectation is according to the joint distribution $p_{Y,U}$. In what follows, we show that (13) can also be efficiently solved through a linear program. We can write

$$\begin{aligned} \mathbb{E}_{U,Y}[(Y-U)^2] &= \mathbb{E}_U\left[\mathbb{E}_{Y|U}[(Y-U)^2|U]\right] \\ &\geq \mathbb{E}_U\left[\mathbb{E}_{Y|U}\left[(Y - \mathbb{E}[Y|U])^2|U\right]\right] \qquad (14) \\ &= \int \mathrm{Var}[Y|U=u]dF_U(u), \qquad (15) \end{aligned}$$

where (14) is a classical result from MMSE estimation [8]. From (10) and (15), we have the following lower bound for (13):

$$\min_{\substack{F_U(\cdot),\ \mathbf{p}_{Y|u}\in\mathbb{S}_\epsilon: \\ \int_{\mathcal{U}} \mathbf{p}_{Y|u} dF(u)=\mathbf{p}_Y}} \int \mathrm{Var}[Y|U=u]dF_U(u), \tag{16}$$

which is tight if and only if $\mathbb{E}[Y|U=u] = u, \forall u \in \mathcal{U}$. It can be verified that $\mathrm{Var}[Y|U=u]$ is a strictly concave function of $\mathbf{p}_{Y|u}$, which follows from the strict convexity of x^2 in x. From the strict concavity of $\mathrm{Var}[Y|U=u]$, we can apply similar reasoning as in the proof of Proposition 2, and conclude that in (16), it is sufficient to consider only the extreme points of $\mathbb{S}_\epsilon$. Therefore, (16) boils down to a standard linear program as follows:

$$\begin{aligned} &\min_{\mathbf{w}\geq 0} \quad \begin{bmatrix} \mathrm{Var}_1 & \mathrm{Var}_2 & \ldots & \mathrm{Var}_K \end{bmatrix} \cdot \mathbf{w}, \\ &\text{s.t. } \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \ldots & \mathbf{p}_K \end{bmatrix} \mathbf{w} = \mathbf{p}_Y \end{aligned} \tag{17}$$

where Var_i ($\forall i \in [1 : K]$) denotes $\mathrm{Var}[Y|U=u]$ under $\mathbf{p}_i$, i.e., when $\mathbf{p}_{Y|u} = \mathbf{p}_i$. Finally, once the LP in (17) is solved, if $w_i^* \neq 0$ ($i \in [1 : K]$), we set $u_i = \mathbb{E}[Y|U=u_i]$, where the expectation is taken under the distribution $\mathbf{p}_{Y|u_i} = \mathbf{p}_i$.

[5] The set $\mathbb{S}_\epsilon$ is an at most $(|\mathcal{Y}|-1)$-dimensional convex subset of $\mathbb{R}^{|\mathcal{Y}|}$. Therefore, any point in $\mathbb{S}_\epsilon$ can be written as a convex combination of at most $|\mathcal{Y}|$ extreme points of $\mathbb{S}_\epsilon$.

[6] Each of these sets correspond to a specific sign determination of the elements in L^1-norm.

C. Minimum probability of error

Another possible utility function is the error probability, which leads to the following trade-off

$$\min_{\substack{p_{U|Y}:\\ X-Y-U\\ T(X;U)\leq\epsilon}} \Pr\{Y\neq U\}. \tag{18}$$

We can write

$$\begin{aligned}\Pr\{Y\neq U\} &= 1-\Pr\{Y=U\}\\ &= 1-\int_{\mathcal{U}}\Pr\{Y=u|U=u\}dF_U(u)\\ &\geq 1-\int_{\mathcal{U}}\max_y p_{Y|U}(y|u)dF_U(u), \qquad (19)\end{aligned}$$

where (19) holds with equality when $u = \arg\max_y p_{Y|U}(y|u)$. Then, (18) is lower bounded by

$$1-\max_{\substack{F_U(\cdot),\ \mathbf{p}_{Y|u}\in\mathbb{S}_\epsilon:\\ \int_{\mathcal{U}}\mathbf{p}_{Y|u}dF(u)=\mathbf{p}_Y}}\int_{\mathcal{U}}\max_y p_{Y|U}(y|u)dF_U(u). \tag{20}$$

It can be verified that $\max_y p_Y(y)$ is convex in $p_Y(\cdot)$. Hence, following a similar reasoning as in the proof of Proposition 2, it is sufficient to consider only the extreme points of $\mathbb{S}_\epsilon$ in the optimization in (20). Therefore, the problem reduces to a standard linear program as follows:

$$\begin{aligned}1-\qquad &\max_{\mathbf{w}\geq 0} \qquad \begin{bmatrix}p_{m_1} & p_{m_2} & \dots & p_{m_K}\end{bmatrix}\cdot\mathbf{w},\\ &\text{s.t. } \begin{bmatrix}\mathbf{p}_1 & \mathbf{p}_2 & \dots & \mathbf{p}_K\end{bmatrix}\mathbf{w}=\mathbf{p}_Y\end{aligned}$$

where p_{m_i} is the maximum element of the vector $\mathbf{p}_i,\ i\in[1:K]$. Once the LP is solved, if $w_i^*\neq 0$ ($i\in[1:K]$), the value of u_i is set as the maximum element of the probability vector $\mathbf{p}_{Y|u_i}=\mathbf{p}_i$.

IV. Motivation of total variation distance as a measure of privacy

The following three subsections motivate the use of total variation distance as a measure of privacy.

A. Post-processing and linkage inequalities

For an arbitrary privacy-leakage measure $J(X;U)$, we have the following definitions from [3].

Definition 1. (***Post-processing inequality***) J satisfies the post-processing inequality if and only if for any Markov chain $A-B-C$, we have $J(A;B)\geq J(A;C)$.

Definition 2. (***Linkage inequality***) J satisfies the linkage inequality if and only if for any Markov chain $A-B-C$, we have $J(B;C)\geq J(A;C)$.

It is obvious that for a symmetric privacy measure, i.e., $J(X;U)=J(U;X)$, like mutual information, the two definitions are equivalent. In [3], the significance of these inequalities for a privacy measure is explained, and it is also shown that some well-known privacy measures, e.g., *differential privacy* or *maximal information leakage*, do not satisfy the linkage inequality.

Theorem 3. The privacy measure $T(\cdot;\cdot)$ given in (1) satisfies both the post-processing and the linkage inequalities.

Proof. Let $A-B-C$ form a Markov chain. We have

$$\begin{aligned}\max_b\|\mathbf{p}_{A|b}-\mathbf{p}_A\|_1 &= \max_{b,c}\|\mathbf{p}_{A|b,c}-\mathbf{p}_A\|_1 \qquad (21)\\ &\geq \max_c\sum_b p_{B|C}(b|c)\|\mathbf{p}_{A|b,c}-\mathbf{p}_A\|_1\\ &\geq \max_c\left\|\sum_b p_{B|C}(b|c)\mathbf{p}_{A|b,c}-\mathbf{p}_A\right\|_1 \qquad (22)\\ &= \max_c\|\mathbf{p}_{A|c}-\mathbf{p}_A\|_1, \qquad (23)\end{aligned}$$

where (21) follows from the fact that $A-B-C$ form a Markov chain; (22) results from the convexity of the L^1-norm. This proves the post-processing inequality.

Let c be an arbitrary realization of the random variable C. We have

$$\begin{aligned}\|\mathbf{p}_{A|c}-\mathbf{p}_A\|_1 &= \|\mathbf{P}_{A|B}(\mathbf{p}_{B|c}-\mathbf{p}_B)\|_1\\ &= \sum_a\left|\sum_b p_{A|B}(a|b)\left(p_{B|C}(b|c)-p_B(b)\right)\right|\\ &\leq \sum_a\sum_b p_{A|B}(a|b)|p_{B|C}(b|c)-p_B(b)| \qquad (24)\\ &= \sum_b\sum_a p_{A|B}(a|b)|p_{B|C}(b|c)-p_B(b)|\\ &= \|\mathbf{p}_{B|c}-\mathbf{p}_B\|_1, \qquad (25)\end{aligned}$$

where (24) follows from the triangle inequality. Taking the maximum over c, this proves that $T(\cdot;\cdot)$, given in (1), satisfies the linkage inequality. □

Remark 1. Among all the L^p-norms ($p\geq 1$), only the L^1-norm satisfies the linkage inequality[7]. Consider the following example: Let $A-B-C$ form a Markov chain, and consider the transition matrix

$$\mathbf{P}_{A|B}=\begin{bmatrix}1 & 1 & \frac{1}{2} & 0 & 0 & 0\\ 0 & 0 & \frac{1}{2} & 1 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 1\end{bmatrix},$$

with $\mathbf{p}_B = \begin{bmatrix}p_1 & p_2 & p_3 & p_4 & p_5 & p_6\end{bmatrix}^T$, $\mathbf{p}_A = \begin{bmatrix}p_7 & p_8 & p_6\end{bmatrix}^T$, where $p_i\in(0,1),\forall i\in[1:8]$ and $\sum_{i=1}^6 p_i = 1$. Let $C\in\{-1,1\}$, and $p_C(1)=\frac{1}{2}$. For sufficiently small $\delta>0$, let $\mathbf{p}_{B|c} = \begin{bmatrix}p_1+c\delta & p_2+c\delta & p_3 & p_4-c\delta & p_5-c\delta & p_6\end{bmatrix}^T$ which results in $\mathbf{p}_{A|c}=\begin{bmatrix}p_7+2c\delta & p_8-2c\delta & p_6\end{bmatrix}^T,\forall c\in\{-1,1\}$. It can be verified that for any $p\in(1,+\infty]$, we have $\|\mathbf{p}_{A|c}-\mathbf{p}_A\|_p > \|\mathbf{p}_{B|c}-\mathbf{p}_B\|_p,\forall c\in\{-1,1\}$.

[7] Note that the quantity $\|\mathbf{x}\|_p = (\sum_i|x_i|^p)^{\frac{1}{p}}$ is not subadditive when $p\in(0,1)$, and thus, does not define a norm. Nonetheless, even if the privacy measure is defined as $J(A;B)=\max_b\|\mathbf{p}_{A|b}-\mathbf{p}_A\|_p$ with $p\in(0,1)$, it can be verified that it does not satisfy the linkage inequality by letting $\mathbf{p}_{B|c}=\begin{bmatrix}p_1 & p_2 & p_3+c\delta & p_4 & p_5 & p_6-c\delta\end{bmatrix}^T,\forall c\in\{-1,1\}$, in the counterexample of this remark.

B. An upper bound on inference threats

An inference threat model is introduced in [1], which models a broad class of statistical inference attacks that can be performed on private data X. Assume that an inference cost function $C(\cdot,\cdot):\mathcal{X}\times\mathcal{P}(\mathcal{X})\to\mathbb{R}$ is given. Prior to observing U, the attacker chooses a belief distribution $\mathbf{q}$ over X as the solution of $c_0^*=\min_{\mathbf{q}\in\mathcal{P}(\mathcal{X})}\mathbb{E}_X[C(X,\mathbf{q})]$, while after observing $U=u$, he revises this belief as the solution of $c_u^*=\min_{\mathbf{q}\in\mathcal{P}(\mathcal{X})}\mathbb{E}_{X|U}[C(X,\mathbf{q})|U=u]$. As a result, the attacker obtains an average gain in inference cost of $\Delta C=c_0^*-\mathbb{E}_U[c_U^*]$, which quantifies the improvement in his inference. A natural way to restrict the attacker's inference quality is to keep ΔC below a target value. The following theorem ensures that for any bounded cost function $C(\cdot,\cdot)$, the attacker's inference quality is restricted in this way by focusing on the control of $T(X;U)$, i.e., keeping it below a certain threshold.

Theorem 4. Let $L=\sup_{x\in\mathcal{X},\mathbf{q}\in\mathcal{P}(\mathcal{X})}|C(x,\mathbf{q})|<+\infty$. We have $\Delta C\leq 4L\cdot T(X;U)$.

Proof. The proof follows similar steps as in [2, Lemma 2] up to the point of using Pinsker inequality. □

Remark 2. It is important to note that the boundedness of the cost function is not a necessary condition. For example, the log-loss cost function, i.e., $C(x,\mathbf{q})=-\log q(x)$, where $q(\cdot)$ is the pmf corresponding to $\mathbf{q}$, is not a bounded cost function. However, ΔC under the log-loss cost function, which is equal to $I(X;U)$, is bounded above by $T(X;U)$ as[8]

$$I(X;U)\leq\log\left(1+\frac{2T^2(X;U)}{\min_x p_X(x)}\right),\tag{26}$$

which is proved by using the reverse Pinsker inequality in [9, Theorem 25].

Remark 3. *Maximal leakage* is proposed as a privacy measure in [4], which is shown to be equivalent to Sibson mutual information of order infinity[9], i.e.,

$$\mathcal{L}(X\to U)=I_\infty(X;U)=\log\sum_{u\in\mathcal{U}}\max_x p_{U|X}(u|x).$$

Maximal leakage is bounded above by $T(X;U)$ as

$$\begin{aligned}\mathcal{L}(X\to U)&=\log\sum_{u\in\mathcal{U}}p_U(u)\max_x\frac{p_{X|U}(x|u)}{p_X(x)}\\&\leq\log\left(1+\frac{T(X;U)}{\min_x p_X(x)}\right),\end{aligned}\tag{27}$$

which follows the fact that for an arbitrary pmf $q_X(\cdot)$,

$$\max_x\frac{q_X(x)}{p_X(x)}\leq\frac{\min_x p_X(x)+\frac{1}{2}\|\mathbf{q}_X-\mathbf{p}_X\|_1}{\min_x p_X(x)}.\tag{28}$$

C. Evaluation of the optimal utility-privacy trade-off

The region $\mathbb{S}_\epsilon$ given in (9) has a simple structure, i.e., a bounded convex set in $\mathcal{P}(\mathcal{Y})$ with a finite number of extreme points, which is a direct consequence of L^1-norm as the privacy measure in (1). This simple structure[10] in turn results in characterizing the optimal utility-privacy trade-off as the solution of a standard linear program as shown in Section III. Other measures of privacy do not necessarily lend themselves to similarly efficient characterization. For example, when mutual information is taken as both the privacy and utility measures, the characterization of the optimal trade-off ($g_\epsilon(X;Y)$ in [10]) is an open problem.

V. Conclusions

We have motivated the employment of the total variation distance as a privacy-leakage measure by showing that i) it satisfies the *post-processing* and *linkage* inequalities; ii) the corresponding optimal utility-privacy trade-off can be efficiently solved through a standard linear program; and iii) it provides a bound on the privacy-leakage measured by the mutual information, the *maximal leakage*, or the improvement in an inference attack with a bounded cost function.

Appendix A

For the set $\mathbb{S}_\epsilon$ defined in (9), let $\mathbf{r}:\mathbb{S}_\epsilon\to\mathbb{R}^{|\mathcal{Y}|}$ be a vector-valued mapping defined element-wise as

$$\begin{aligned}r_i(p_{Y|U}(\cdot|u))&=p_{Y|U}(y_i|u),\ i\in[1:|\mathcal{Y}|-1]\\r_{|\mathcal{Y}|}(p_{Y|U}(\cdot|u))&=H(Y|U=u).\end{aligned}$$

Since $\mathbb{S}_\epsilon$ is a closed and bounded subset of $\mathbb{R}^{|\mathcal{Y}|}$, it is compact. Also, $\mathbf{r}$ is a continuous mapping from $\mathbb{S}_\epsilon$ to $\mathbb{R}^{|\mathcal{Y}|}$. From the support lemma [11], the bound on $|\mathcal{U}|$ can be obtained [5].

References

[1] F. Calmon and N. Fawaz, "Privacy against statistical inference," in *50th Annual Allerton Conference*, Illinois, USA, Oct. 2012, pp. 1401–1407.

[2] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Médard, "From the information bottleneck to the privacy funnel," in *IEEE Information Theory Workshop (ITW)*, 2014, pp. 501–505.

[3] Y. Wang, Y. Basciftci, and P. Ishwar, "Privacy-utility tradeoffs under constrained data release mechanisms," *https://arxiv.org/pdf/1710.09295.pdf*, Oct. 2017.

[4] I. Issa, S. Kamath, and A. Wagner, "An operational measure of information leakage," in *Information Science and Systems (CISS)*, 2016, pp. 234–239.

[5] B. Rassouli and D. Gündüz, "Optimal utility-privacy trade-off with total variation distance as a privacy measure," *Available online at arXiv:1801.02505*.

[6] D. Bertsimas and J. N. Tsitsiklis, *Introduction to linear optimization*. Athena Scientic, 1997.

[7] K. G. Murty, *Linear Programming*. John Wiley and Sons, 1983.

[8] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. Springer, 2008.

[9] I. Sason and S. Verdú, "f -divergence inequalities," *IEEE Trans. Inf. Theory*, pp. 5973–6006, 2016.

[10] S. Asoodeh, F. Alajaji, and T. Linder, "Notes on information-theoretic privacy," in *52nd Annual Allerton Conference*, Illinois, USA, Oct. 2014, pp. 1272–1278.

[11] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2012.

[8] When $T(X;U)>\frac{1}{2}$, (27) gives a tighter upper bound, which follows from the fact that $I(X;U)\leq\mathcal{L}(X\to U)$ in [4].

[9] Note that we have $p_X(x)>0,\forall x\in\mathcal{X}$.

[10] This is also the case in the more general observation model in [3], i.e., for the Markov chain $(X,Y)-W-U$, it is sufficient to consider the inverse image of $\tilde{\mathbb{S}}_\epsilon$ in (8) under $\mathbf{P}_{X|W}$.