Tunable Measures for Information Leakage and Applications to Privacy-Utility Tradeoffs

Jiachun Liao, *Student Member, IEEE*, Oliver Kosut, *Member, IEEE*,
Lalitha Sankar, *Senior Member, IEEE*, and Flavio du Pin Calmon, *Member, IEEE*

***Abstract*—We introduce a tunable measure for information leakage called *maximal α-leakage*. This measure quantifies the maximal gain of an adversary in inferring any (potentially random) function of a dataset from a release of the data. The inferential capability of the adversary is, in turn, quantified by a class of adversarial loss functions that we introduce as α-loss, $\alpha \in [1,\infty) \cup \{\infty\}$. The choice of α determines the specific adversarial action and ranges from refining a belief (about any function of the data) for $\alpha = 1$ to guessing the most likely value for $\alpha = \infty$ while refining the α^{th} moment of the belief for α in between. Maximal α-leakage then quantifies the adversarial gain under α-loss over all possible functions of the data. In particular, for the extremal values of $\alpha = 1$ and $\alpha = \infty$, maximal α-leakage simplifies to mutual information and maximal leakage, respectively. For $\alpha \in (1,\infty)$ this measure is shown to be the Arimoto channel capacity of order α. We show that maximal α-leakage satisfies data processing inequalities and a sub-additivity property thereby allowing for a weak composition result. Building upon these properties, we use maximal α-leakage as the privacy measure and study the problem of data publishing with privacy guarantees, wherein the utility of the released data is ensured via a *hard distortion* constraint. Unlike average distortion, hard distortion provides a deterministic guarantee of fidelity. We show that under a hard distortion constraint, for $\alpha > 1$ the optimal mechanism is independent of α, and therefore, the resulting optimal tradeoff is the same for all values of $\alpha > 1$. Finally, the tunability of maximal α-leakage as a privacy measure is also illustrated for binary data with average Hamming distortion as the utility measure.**

***Index Terms*—Mutual information, maximal leakage, maximal α-leakage, Sibson mutual information, Arimoto mutual information, f-divergence, privacy-utility tradeoff, hard distortion.**

I. Introduction and Overview

The measure and control of private information leakage is a recognized objective in communications, information theory, and computer science. Modern cryptography [1]–[3], for example, aims at designing and analyzing security systems that are believed to be impervious to computationally bounded adversaries. Alternatively, information-theoretic security studies settings where an asymmetry of information between an adversary and the legitimate parties (e.g., the wiretap channel [4]–[6]) can be exploited to guarantee that no private information is leaked regardless of computational assumptions. An adversary that *only* observes the output of a (computationally) secure cipher or cannot overcome the information asymmetry in a wiretap-like setting does not, for all practical purposes, pose a privacy risk.

However, modern applications such as online data sharing, social networks, cloud-based services, and mobile computing have significantly increased the number of ways in which private information can leak. Services that require a user to disclose data in order to receive utility inevitably incur a privacy risk through unwanted inferences. For example, *sensitive information* such as political preference, medical conditions, and identity can be reliably estimated from movie ratings [7], online shopping patterns, [8], and via deanonymization and tracking of interactions in social network data [9], [10], respectively. Moreover, practical implementations of cryptographic schemes are susceptible to so-called "side-channel attacks," where sensitive information leaks through unexpected channels. For example, a malicious application may get timing characteristics [11], [12]. In these examples, an adversary that observes information leaked through a side-channel can more reliably infer private data, such as a key or a plaintext.

Several (often overlapping) definitions of privacy/information leakage have been proposed over the past decade. The most widely adopted measure is differential privacy (DP) [13], [14], which was introduced within the context of querying databases. DP seeks to ensure that changes in the database entries do not significantly influence the value of a query. A variety of information-theoretic measures have also been proposed as leakage measures. Foremost among them is mutual information (MI): its use as a privacy measure in [15]–[24] is inspired by the common appearance of MI as an operationally-meaningful quantity throughout the literature on communication systems. In a similar vein, divergence-based quantities such as total variation distance between the prior and posterior distributions [25] have also been proposed as leakage measures. Information-theoretic measures have been studied in the DP community via Rényi differential privacy which is based on Rényi divergence [26] that allow relaxing the original definition of DP in order to enable better utility guarantees. However, the gamut of information-theoretic leakage measures proposed to address the privacy problem do not *yet* have clear operational meanings or adversarial models in their definitions.

More recently, information-theoretic formulations have been introduced to capture privacy against a "guessing" adversary. Here, privacy is measured in terms of an adversary's gain in guessing the private information after observing disclosed data. For example, Asoodeh et al. use the probability of correctly guessing to measure privacy [27]; and Issa et al. introduce

This material is based upon work supported by the National Science Foundation under Grant Nos. CCF-1422358, CCF-1350914, and CIF-1422358. This work was presented in part at IEEE International Symposium on Information Theory and Information Theory Workshop in 2018.

maximal leakage (MaxL), which quantifies the maximal logarithmic gain in the probability of correctly guessing any arbitrary function of the original data from released data [28]. A related line of work includes [29]–[31], where security is quantified in terms of the expected number of guesses (or moments thereof) required by an adversary to correctly identify a quantity of interest (e.g., a password or a transmitted codeword).

This work builds upon the abovementioned efforts to operationally motivate measures and presents a larger class of meaningful information-theoretic measures that can be operationally motivated in the privacy setting. To this end, we introduce a tunable loss function, namely α-loss ($1 \leq \alpha \leq \infty$), to capture adversarial actions. In particular, for $\alpha = 1$ and $\alpha = \infty$ the loss function simplifies to the logarithmic loss (log-loss) [32]–[34] and the probability of error[1], respectively. The choice of the loss function captures the *inferential action* of an adversary. Specifically, the adversarial action, henceforth referred to as inference, involves refining a posterior belief of one or more sensitive features. Adversarial gain of a computationally unbounded adversary is then simply the decrease in (inferential) loss on average as a result of a data release.

We use the α-loss function to derive two new privacy measures called *α-leakage* and *maximal α-leakage*. Specifically, α-leakage quantifies an adversary's gain in inferring a *specific* private attribute in the dataset; in contrast, maximal α-leakage quantifies an adversary's gain in inferring *any arbitrary* attribute of the dataset. In particular, maximal α-leakage includes MI and MaxL as special cases for $\alpha = 1$ and $\alpha = \infty$, respectively. This approach allows us to show that MaxL can be interpreted in terms of an adversary seeking to minimize the 0-1 loss function [33], [35] ($\alpha = \infty$), i.e., the adversary makes a hard decision via a maximum likelihood estimator. On the other hand, we show that when MI is used as a leakage measure ($\alpha = 1$), the underlying loss function is the log-loss, that models a (soft decision) belief-refining adversary. In addition to what the adversary observes (e.g., released census dataset or information via a side-channel), the adversary may also have access to other correlated side-information (e.g., voter record database or individual personal information in side-channel attacks); generalizing *α-leakage* and *maximal α-leakage* to model such side-information is indeed possible as recently shown by the authors in [36]; however, this generalization is beyond the scope of this paper.

Our proposed measures can be applied to the aforementioned privacy and side-channel settings. In most non-trivial settings of data publishing, there is a fundamental privacy-utility tradeoff (PUT): on the one hand, releasing data "as is" can lead to unwanted inferences of private information. On the other hand, perturbing or limiting the released data reduces its quality. We quantify PUTs for two types of data models: one in which the entire dataset is sensitive (as illustrated in Fig. 1a) and the other in which only a subset of the dataset is sensitive (as illustrated in Fig. 1b). Throughout this paper, we use X to denote the original data that will be *released* as Y via a randomized mapping; X may be entirely sensitive as in Fig. 1a, or it may be separate from the sensitive features S as in Fig. 1b. The variable U represents a specific sensitive feature of the dataset that the adversary is interested in learning. Examples of datasets wherein the entire data is sensitive include data collected by smart devices such as smartphone sensors, movie recommendation systems, where it is hard to know *a priori* which aspect of the data ought to be identified as sensitive. In contrast, examples of datasets with clearly defined sensitive features include census and other datasets that explicitly include personally identifiable information.

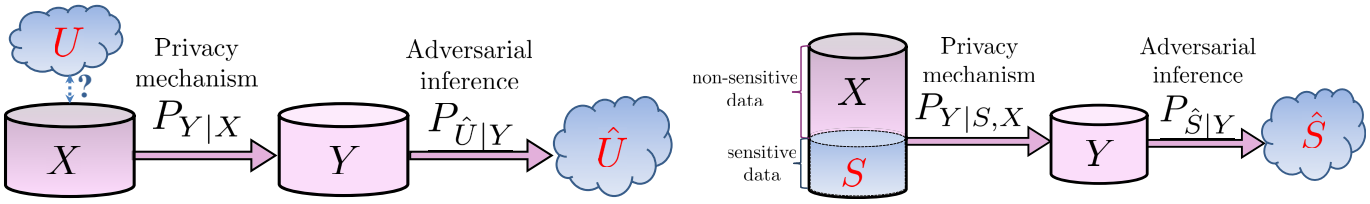
The exact nature of the PUT depends on exactly how both privacy and utility are measured. Towards an understanding of our new privacy measures, we consider PUTs in which (maximal) α-leakage is the privacy measure, and we study several options for utility measure. In general, a meaningful utility measure (between the original and released data) should require the released data to provide either (i) average-case guarantees on fidelity [18], [25], [27], [37], [38]; or (ii) worst-case guarantees on fidelity. Indeed, requirement (i) lends itself to modeling with a large class of expected value constraints including average distortion constraints and is now well studied in information-theoretical privacy via a variety of measures such as Hamming distortion, square error and Kullback–Leibler divergence [23], [38]–[41]. We note that average distortion constraints are also well studied in rate-distortion theory. To capture utility requirement (ii), we introduce a *hard distortion* measure which constrains the privacy mechanism so that the distortion between original and released datasets is bounded with probability 1. Such an approach has also been studied in rate-distortion theory as a potential distortion measure (see, for example, [42] for the use of per symbol distortion constraints). In addition, compared to average-case distortion constraints [23], [38]–[41], a hard distortion measure is quite stringent but allows the data curator to make specific, deterministic guarantees on the fidelity of the released dataset relative to the original. Such a deterministic guarantee can lead to more accurate statistical estimators, e.g., the empirical distribution estimation for publicly released datasets such as the census.

A. Contributions and Organization

The main contributions of this paper include:

- We introduce a tunable loss function, namely α-loss ($1 \leq \alpha \leq \infty$), which captures log-loss and 0-1 loss, respectively, for extremal values of $\alpha = 1$ and ∞, respectively (Sec. III-A).
- Based on α-loss, we define two operational measures of information leakage: α-leakage and maximal α-leakage, and show that: (i) α-leakage equals to Arimoto mutual information of order α [43], [44]; and (ii) maximal α-leakage equals to MI for $\alpha = 1$ and Arimoto channel capacity [44] of order α for $\alpha > 1$. Note that maximal α-leakage captures MI and MaxL at the extremal values of α (Sec. III-B). The proofs of these results rely on the fact that maximizing either the Arimoto MI or the Sibson MI [45] over the input distribution yields the same quantity, the Arimoto channel capacity.

[1]Note that the probability of error for a maximum likelihood estimator is exactly the 0-1 loss [33], [35].



(a) The privacy protection for entirely sensitive datasets.

(b) The privacy protection for datasets with non-sensitive and sensitive data.

Fig. 1: Two privacy-guaranteed data publishing scenarios: (i) the left figure shows the privacy protection for entirely sensitive datasets, where X and Y represent the original and released data. An adversary intends to infer a function U of X from Y, and $\hat{U}$ is the adversary's estimation of U. Generally, the function U is unknown to the data curator/provider; (ii) the right figure shows the privacy protection for datasets consisting of non-sensitive and sensitive data, where X and S represent the non-sensitive and sensitive data in original dataset, respectively, and Y is the released version of X. The adversary intends to infer S from Y, and $\hat{S}$ is the adversary's estimation of S.

- Inspired by the fact that maximal α-leakage equals to the Arimoto channel capacity, we introduce a broader class of information-leakage measures based on f-divergences, which capture maximal α-leakage as a special case (Sec. III-C);
- We prove that maximal α-leakage satisfies several useful properties, including: (i) quasi-convexity, (ii) data-processing inequalities: post-processing inequality and linkage inequality, (iii) sub-additivity (iv) additivity for memoryless mappings (Sec. IV).
- In the context of privacy-guaranteed data publishing subject to a hard distortion utility constraint on data, we solve the resulting PUT problems exactly for maximal α-leakage as well as its f-divergence-based variants (Sec. V-A). For α-leakage, which restricts leakage about specific sensitive data as shown in Fig. 1b, we provide an inner bound of the optimal PUT (Sec. V-B). In Sec. VI, we illustrate these results via two examples.

II. Preliminaries

We use capital letters to represent *discrete* random variables, and the corresponding capital calligraphic and lower-case letters represent their *finite* supports and the elements of the supports, respectively. For example, for a random variable X, its support is $\mathcal{X}$ with any possible realization $x \in \mathcal{X}$. In addition, we use $\log$ to represent the natural logarithm, and $[a,b]$ to indicate the set of integers from a to b. We use $|\cdot|$ to indicate the cardinality of a set, e.g., $|\mathcal{X}|$, and $\|\cdot\|_p$ to represent the p-norm of a vector, e.g., for $\alpha \geq 1$, $\|P_X\|_\alpha \triangleq (\sum_{x\in\mathcal{X}} P_X(x)^\alpha)^{\frac{1}{\alpha}}$.

We begin by reviewing Rényi entropy and divergence [46], [47].

Definition 1. *Given a distribution P_X, the Rényi entropy of order $\alpha \in (0,1)\cup(1,\infty)$ is defined as*

$$H_\alpha(P_X) = \frac{1}{1-\alpha}\log\sum_{x\in\mathcal{X}} P_X(x)^\alpha, \tag{1}$$

$$= \frac{\alpha}{1-\alpha}\log\|P_X\|_\alpha, \quad (\alpha \geq 1). \tag{2}$$

Let Q_X be a distribution over the support of P_X. The Rényi divergence (between P_X and Q_X) of order $\alpha \in (0,1)\cup(1,\infty)$ is defined as

$$D_\alpha(P_X\|Q_X) = \frac{1}{\alpha-1}\log\left(\sum_{x\in\mathcal{X}}\frac{P_X(x)^\alpha}{Q_X(x)^{\alpha-1}}\right). \tag{3}$$

Both of the two quantities are defined by their continuous extensions for $\alpha = 1$ and ∞. Specifically, for $\alpha = \infty$, the two quantities are given by

$$H_\infty(P_X) = \min_x \log\frac{1}{P_X(x)}, \tag{4}$$

which is called min-entropy, and

$$D_\infty(P_X\|Q_X) = \log\max_x \frac{P_X(x)}{Q_X(x)}. \tag{5}$$

For $\alpha = 1$, the Rényi entropy and divergence reduce to Shannon entropy and Kullback-Leibler divergence, respectively [43].

The α-leakage and maximal α-leakage measures can be expressed in terms of Sibson MI [45] and Arimoto MI [44]. These quantities generalize the usual notion of MI. We review these definitions next.

Definition 2. *Let discrete random variables $(X,Y) \sim P_{X,Y}$ with P_X and $P_{Y|X}$ as the marginal and conditional distributions, respectively, and Q_Y be an arbitrary distribution over the finite support $\mathcal{Y}$. The Sibson mutual information of order $\alpha \in (0,1)\cup(1,\infty)$ is defined as*

$$I_\alpha^S(X;Y) \triangleq \inf_{Q_Y} D_\alpha(P_{X,Y}\|P_X\times Q_Y) \tag{6}$$

$$= \frac{\alpha}{\alpha-1}\log\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P_X(x)P_{Y|X}(y|x)^\alpha\right)^{\frac{1}{\alpha}}. \tag{7}$$

The Arimoto mutual information of order $\alpha \in (0,1)\cup(1,\infty)$ is defined as

$$I_\alpha^A(X;Y) \triangleq H_\alpha(X) - H_\alpha^A(X|Y) \tag{8}$$

$$= \frac{\alpha}{\alpha-1}\log\frac{\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P_{X,Y}(x,y)^\alpha\right)^{\frac{1}{\alpha}}}{\left(\sum_{x\in\mathcal{X}} P_X(x)^\alpha\right)^{\frac{1}{\alpha}}}, \tag{9}$$

$$= \frac{\alpha}{\alpha-1} \log \frac{\sum\limits_{y\in\mathcal{Y}} \|P_{X,Y}(\cdot,y)\|_\alpha}{\|P_X\|_\alpha}, \quad (\alpha \geq 1) \tag{10}$$

where $H_\alpha^A(X|Y)$ is Arimoto conditional entropy of X given Y defined as

$$H_\alpha^A(X|Y) = \frac{\alpha}{1-\alpha} \log \sum_{y\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}} P_{X,Y}(x,y)^\alpha \right)^{\frac{1}{\alpha}}. \tag{11}$$

All of these quantities are defined by their continuous extension for $\alpha = 1$ or ∞.

Note that for $\alpha = 1$, both Sibson and Arimoto MIs reduce to Shannon's MI; however, for $\alpha = \infty$, the Sibson MI is

$$I_\infty^{\mathrm{S}}(X;Y) = \log \sum_y \max_x P_{Y|X}(y|x), \tag{12}$$

and the Arimoto MI is given by

$$I_\infty^{\mathrm{A}}(X;Y) = \log \frac{\sum\limits_y \max\limits_x P_{X,Y}(x,y)}{\max\limits_x P_X(x)}. \tag{13}$$

The two measures of information generalize Shannon's MI and have a number of interesting and useful properties in various problems [43]–[45], [48].

III. Tunable Loss Function and Information Leakage Measures

Information leakage of a data release can be viewed as an increase in adversarial inference as a result of the data release. This inference performance can be precisely characterized by a loss function that an adversary minimizes. In this section, we introduce a tunable loss function, namely *α-loss* for $\alpha \in [1,\infty]$, to captures a computationally unbounded adversary's inference in refining a posterior belief of one or more sensitive features from a data release, and introduce two tunable measures, called *α-leakage* and *maximal α-leakage*, respectively, to measure the corresponding information leakages due to the data release.

A. α-Loss Function

For a Markov chain $X - Y - \hat{X}$, let $\hat{X}$ be an estimator of X and $P_{\hat{X}|Y}$ be a strategy for estimating X from Y. We denote the probability of correctly estimating $X = x$ given $Y = y$ as $P_{\hat{X}|Y}(x|y)$. The estimation strategy $P_{\hat{X}|Y}$ is selected in order to minimize an *expected loss* measure. Denoting the loss function by $\ell(x, y, P_{\hat{X}|Y})$, the expected loss is given by $\mathbb{E}\big[\ell\big(X, Y, P_{\hat{X}|Y}\big)\big]$.

One formulation of the loss function is the probability of *incorrectly* guessing given by

$$\ell_{0-1}(x, y, P_{\hat{X}|Y}) = 1 - P_{\hat{X}|Y}(x|y), \tag{14}$$

such that the expected loss $\mathbb{E}\big[\ell_{0-1}\big(X, Y, P_{\hat{X}|Y}\big)\big]$ is the expected probability of error. Here, the optimal strategy $P^\star_{\hat{X}|Y}$ is the standard maximal posterior (MAP) estimator given by

$$P^\star_{\hat{X}|Y}(x|y) = \begin{cases} 1, & x = \arg\max\limits_{x\in\mathcal{X}} P_{X|Y}(x|y) \\ 0, & \text{otherwise} \end{cases}, \tag{15}$$

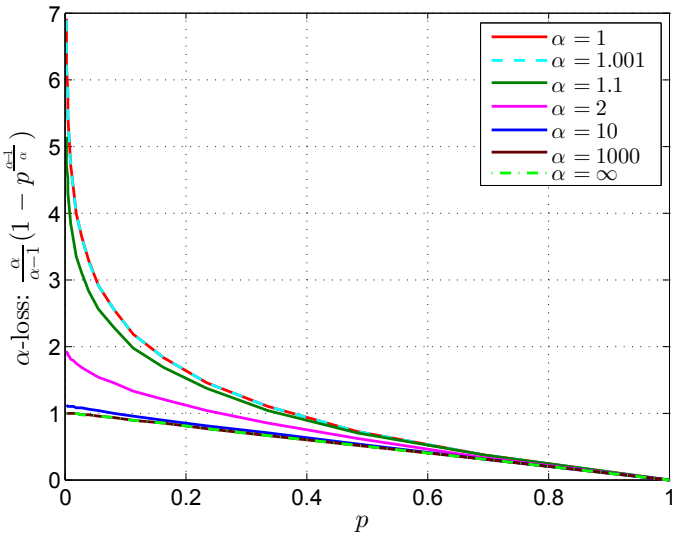


Fig. 2: The plot of α-loss as a function of p. Note that the $p \in [0.001, 1]$ represents the probability of correctly guessing, i.e., $p = P_{\hat{X}|Y}(x|y)$ with an observation $Y = y$ and $p = P_{\hat{X}}(x)$ without any observation.

which makes the loss $\ell_{0-1}(x, y, P^\star_{\hat{X}|Y})$ be either 0 or 1, and therefore, called 0-1 *loss* in the literature [33], [35], [49]. The corresponding expected loss $\mathbb{E}\big[\ell_{0-1}\big(X, Y, P^\star_{\hat{X}|Y}\big)\big]$ is the minimal expected probability of error.

To measure the uncertainty for the strategy $P_{\hat{X}|Y}$, the log-loss (used, for example, in [32]–[34], [50]) is given by

$$\ell_{\log}(x, y, P_{\hat{X}|Y}) = \log \frac{1}{P_{\hat{X}|Y}(x|y)}. \tag{16}$$

The expected loss in this case is the conditional cross-entropy, given by

$$\begin{aligned} &\mathbb{E}\left[\ell_{\log}(X, Y, P_{\hat{X}|Y})\right] \\ =&\sum_{x,y} P_{X,Y}(x,y) \log \frac{1}{P_{\hat{X}|Y}(x|y)}, &(17)\\ =&H(X|Y) + \sum_y P_Y(y) D(P_{X|Y=y} \| P_{\hat{X}|Y=y}). &(18) \end{aligned}$$

Therefore, the optimal strategy is the true posterior distribution of X given Y, i.e., $P^\star_{\hat{X}|Y} = P_{X|Y}$, which makes the expected loss in (18) become the conditional entropy $H(X|Y)$. That is, the minimal expected log-loss is the true conditional entropy.

Note that both the 0-1 loss and log-loss functions are decreasing in the probability of correctly estimation $P_{\hat{X}|Y}(x|y)$. Specifically, for $P_{\hat{X}|Y}(x|y) = 1$, both the values of 0-1 loss and α-loss are 0, and for $P_{\hat{X}|Y}(x|y) = 0$, the values of 0-1 loss and log-loss become 1 and ∞, respectively. To allow a continuous quantification of the loss for $P_{\hat{X}|Y}(x|y) = 0$ from 1 to ∞, we formally define a tunable loss function, namely *α-loss*, as follows.

Definition 3 (α-loss)**.** *Let random variables X, Y and $\hat{X}$ form a Markov chain $X - Y - \hat{X}$, where $\hat{X}$ is an estimator of X.*

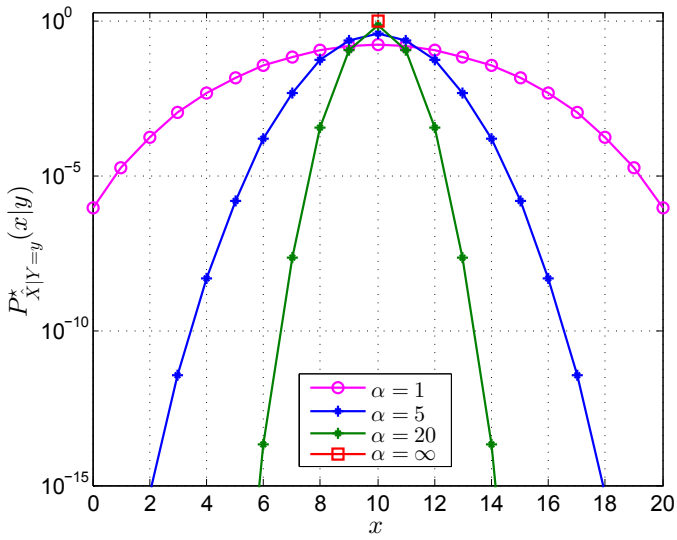


Fig. 3: The optimal strategy in (23) for different α. Note that the magenta circles represent the true conditional probability $P_{X|Y=y}$, which is a binomial distribution with parameters $(n, p) = (20, 0.5)$.

The α-loss of the strategy $P_{\hat{X}|Y}$ for estimating X from Y is

$$\ell_\alpha(x, y, P_{\hat{X}|Y}) = \frac{\alpha}{\alpha - 1}\left(1 - P_{\hat{X}|Y}(x|y)^{\frac{\alpha-1}{\alpha}}\right), \tag{19}$$

where $\alpha \in (1, \infty)$. It is defined by its continuous extension for $\alpha = 1$ and $\alpha = \infty$, respectively, and is given by

$$\ell_1(x, y, P_{\hat{X}|Y}) = \lim_{\alpha\to 1} \ell_\alpha(x, y, P_{\hat{X}|Y}) = \log \frac{1}{P_{\hat{X}|Y}(x|y)}, \tag{20}$$

$$\ell_\infty(x, y, P_{\hat{X}|Y}) = \lim_{\alpha\to\infty} \ell_\alpha(x, y, P_{\hat{X}|Y}) = 1 - P_{\hat{X}|Y}(x|y). \tag{21}$$

Note that for $\alpha = 1$, the expression in (20) follows directly from the L'Hôpital's rule and α-loss becomes the log-loss in (16); and for $\alpha = \infty$, the loss in (21) is exactly the probability of error in (14), which becomes 0-1 loss for MAP estimators. Fig. 2 plots the α-loss function in (19) for different values of α. From Fig. 2, we observe that α-loss function is decreasing and convex in the probability of correctly guessing.

Lemma 1. *For $1 \le \alpha \le \infty$, the minimal expected α-loss is given by*

$$\min_{P_{\hat{X}|Y}} \mathbb{E}\left[\ell_\alpha(X, Y, P_{\hat{X}|Y})\right] = \begin{cases} \frac{\alpha}{\alpha-1}\left(1 - \exp\left(\frac{1-\alpha}{\alpha} H^A_\alpha(X|Y)\right)\right), & \alpha > 1 \\ H(X|Y), & \alpha = 1 \end{cases}, \tag{22}$$

with the optimal estimation strategy given by [2]

$$P^\star_{\hat{X}|Y}(x|y) = \frac{P_{\hat{X}|Y}(x|y)^\alpha}{\sum_{x\in\mathcal{X}} P_{\hat{X}|Y}(x|y)^\alpha}. \tag{23}$$

A detailed proof is in Appendix A. Note that in (22), $H^{\mathrm{A}}_\alpha(X|Y)$ is Arimoto conditional entropy of X given Y in (11). For $\alpha = \infty$, the expression of $H^{\mathrm{A}}_\infty(X|Y)$ is

$$H^{\mathrm{A}}_\infty(X|Y) = \log \sum_y P_Y(y) \max_x P_{X|Y}(x|y), \tag{24}$$

such that $\exp\left(H^{\mathrm{A}}_\infty(X|Y)\right)$ is the maximal expected probability of correctly guessing X from Y. Therefore, for $\alpha = \infty$, the minimal expected α-loss is the minimal expected probability of error. In addition, the optimal estimation strategy in (23) becomes the true posterior distribution of X for $\alpha = 1$ and the MAP estimator for $\alpha = \infty$, respectively.

Example 1. *Let the conditional probability distribution of X given $Y = y$ be a binomial distribution with parameters $(n, p) = (20, 0.5)$, i.e., $P_{X|Y}(x|y) = \binom{20}{x} 0.5^x 0.5^{20-x}$ for $x \in [0, 20]$. Fig. 3 shows the optimal strategies in (23) for different values of α. We observe from Fig. 3 that as α grows from 1 to ∞, the optimal strategy gradually eliminates the less likely values of X (given y) and transforms from the true posterior distribution to the MAP estimator.*

B. α-Leakage and Maximal α-Leakage

Let X and Y represent the original data and released data, respectively, and let U represent an arbitrary (potentially random) function of X that the observer (a curious or malicious user of the released data Y) is interested in learning. In [28], Issa *et al.* introduced MaxL to quantify the maximal gain in an adversary's ability of guessing U after observing Y. We review the definition below.

Definition 4 ([28, Def. 1])**.** *Given a joint distribution $P_{X,Y}$ on finite alphabets, the maximal leakage from X to Y is*

$$\mathcal{L}_{MaxL}(X \to Y) \triangleq \sup_{U-X-Y} \log \frac{\max_{P_{\hat{U}|Y}} \mathbb{E}\left[P_{\hat{U}|Y}(U|Y)\right]}{\max_u P_U(u)}, \tag{25}$$

where $\hat{U}$ represents an estimator taking values from the same arbitrary finite support as U.

Note that the numerator of the logarithmic term in (25) is the maximal expected probability of correctly guessing U with Y given by

$$\max_{P_{\hat{U}|Y}} \mathbb{E}\left[P_{\hat{U}|Y}(U|Y)\right] = \max_u \sum_y P_Y(y) P_{U|Y}(u|y), \tag{26}$$

which is exactly the complement of the minimal expected 0-1 loss in guessing U with Y. Similarly, the denominator is the complement of the minimal expected 0-1 loss in guessing U without Y. Therefore, MaxL is a leakage measure related to 0-1 loss in (14).

In addition, in Def. 4, U represents any (possibly random) function of X. The numerator represents the maximal probability of correctly guessing U based on Y, while the denominator represents the maximal probability of correctly guessing U *without* knowing Y. Thus, MaxL quantifies the maximal logarithmic gain in guessing any possible function of X when an adversary has access to Y.

Analogously to the derivation of MaxL from 0-1 loss, we introduce α-leakage and maximal α-leakage based on α-loss

[2]Note that if there are more than one realization sharing the same maximal posterior belief, for $\alpha = \infty$ the optimal strategy in (23) will output these most likely values with the same probability.

(under the assumptions of discrete random variables and finite supports). The formal definitions are as follows.

Definition 5 (α-Leakage)**.** *Given a joint distribution $P_{X,Y}$ and an estimator $\hat{X}$ with the same support as X, the α-leakage from X to Y is defined as*

$$\mathcal{L}_\alpha(X \to Y) \triangleq \frac{\alpha}{\alpha-1} \log \frac{\max\limits_{P_{\hat{X}|Y}} \mathbb{E}\left[P_{\hat{X}|Y}(X|Y)^{\frac{\alpha-1}{\alpha}}\right]}{\max\limits_{P_{\hat{X}}} \mathbb{E}\left[P_{\hat{X}}(X)^{\frac{\alpha-1}{\alpha}}\right]}, \quad (27)$$

for $\alpha \in (1,\infty)$ and by the continuous extension of (27) for $\alpha = 1$ and ∞.

Note that for any specific function U of X, the joint probability distribution of X and the U is known, and therefore, α-leakage can also be used to measure the the inference gain in inferring the specific function U from the released data Y. In addition, the two maximizations in the numerator and denominator of the logarithmic ratio in (27) imply the optimal adversarial actions in the sense of minimizing the expected α-loss in Lemma 1. Therefore, it limits the inference gain that an adversary can obtain by minimizing the expected α-loss, no matter the adversary has prior knowledge (i.e., the probability distribution of the original data) of the original data or not.

Whereas α-leakage captures how much an adversary can learn about X (or a specific function of X) from Y, we also wish to quantify the information leaked about *any function* of X through Y. To this end, we define maximal α-leakage below.

Definition 6 (Maximal α-Leakage)**.** *Given a joint distribution $P_{X,Y}$ on finite alphabets $\mathcal{X}\times\mathcal{Y}$, the maximal α-leakage from X to Y is defined as*

$$\mathcal{L}_\alpha^{max}(X \to Y) \triangleq \sup_{U-X-Y} \mathcal{L}_\alpha(U;Y), \quad (28)$$

where $1 \le \alpha \le \infty$, and U represents any function of X and takes values from an arbitrary finite alphabet.

Note that for $\alpha \ge 1$,

$$\begin{aligned}&\max_{P_{\hat{U}|Y}} \mathbb{E}\left[P_{\hat{U}|Y}(U|Y)^{\frac{\alpha-1}{\alpha}}\right]\\ =&1 - \frac{\alpha-1}{\alpha} \min_{P_{\hat{U}|Y}} \mathbb{E}\left[\ell_\alpha(U,Y,P_{\hat{U}|Y})\right]. \end{aligned} \quad (29)$$

Thus, there is a similar connection between maximal α-leakage and α-loss (in Def. 3) as that observed in (26) between MaxL and 0-1 loss, and maximal α-leakage quantifies an adversary's capability to infer *any function* of data X from the released Y.

Making use of the result in Lemma 1, the following theorem simplifies the expression of α-leakage in (27).

Theorem 1. *For $1 \le \alpha \le \infty$, α-leakage defined in (27) simplifies to*

$$\mathcal{L}_\alpha(X \to Y) = I_\alpha^A(X;Y). \quad (30)$$

From (29) and Lemma 1, we simplify the scaled logarithm of the ratio in (27) to Arimoto MI. A detailed proof is in Appendix B, where we show that Arimoto conditional entropy and Rényi entropy capture the inference uncertainties of an adversary for knowing Y or not, respectively, and α-leakage measures the decrease in the inference uncertainty by knowing Y.

Making use of the conclusion in Thm. 1, the following theorem gives equivalent expressions for maximal α-leakage. Note that in the following theorem we use the well-known equivalence of the supremums of Sibson and Arimoto MIs [43, Thm. 5].

Theorem 2. *For $1 \le \alpha \le \infty$, the maximal α-leakage defined in (28) simplifies to*

$$\begin{aligned}&\mathcal{L}_\alpha^{max}(X \to Y)\\ =&\begin{cases} \sup\limits_{P_{\tilde{X}}} I_\alpha^S(\tilde{X};Y) = \sup\limits_{P_{\tilde{X}}} I_\alpha^A(\tilde{X};Y), & 1 < \alpha \le \infty \quad (31a)\\ I(X;Y), & \alpha = 1 \quad (31b)\end{cases}\end{aligned}$$

where $P_{\tilde{X}}$ is a probability distribution over the support of P_X.

Note that maximal α-leakage is essentially the Arimoto channel capacity (with a support-set constrained input distribution) for $\alpha > 1$ [44], which is used to characterize probabilities of decoding error for scenarios in which transmission rates are higher than channel capacity. The limit of maximal α-leakage for $\alpha = 1$ gives the Shannon channel capacity. Recall that the limit of α-loss in (19) leads to the log-loss (for $\alpha = 1$) and 0-1 loss (for $\alpha = \infty$) functions, respectively. Consequently, for $\alpha = 1$ and ∞, maximal α-leakage simplifies to MI and MaxL, respectively.

A detailed proof for Thm. 2 is in Appendix C. We summarize key steps in the proof as follows: by applying Thm. 1, we write maximal α-leakage as

$$\mathcal{L}_\alpha^{\max}(X \to Y) = \sup_{U-X-Y} I_\alpha^{\mathrm{A}}(U;Y) \quad \alpha \in [1,\infty]. \quad (32)$$

For $\alpha = 1$, Arimoto MI is simply the Shannon's MI, and combining with the data processing inequalities, (32) simplifies to $I(X;Y)$. Note that for $\alpha > 1$, Arimoto MI does not satisfy data processing inequalities. By using the facts that Arimoto MI and Sibson MI have the same supremum [43, Thm. 5] and that Sibson MI satisfies data processing inequalities [43, Thm. 3], we limit the supremum in (32) by $\sup_{P_{\tilde{X}}} I_\alpha^{\mathrm{S}}(\tilde{X};Y)$, and then, show that the upper bound $\sup_{P_{\tilde{X}}} I_\alpha^{\mathrm{S}}(\tilde{X};Y)$ can be achieved by a specific U with $H(X|U) = 0$.

Example 2. *Given a binary channel*

$$P_{Y|X} = \begin{bmatrix} 1-\rho_1 & \rho_1 \\ \rho_2 & 1-\rho_2 \end{bmatrix}, \quad (33)$$

where $\rho_1, \rho_2 \in [0,1]$ are the crossover probabilities, maximal α-leakage in (31) is given by

$$\begin{aligned}&\mathcal{L}_\alpha^{max}(X \to Y)\\ =&\frac{\alpha}{\alpha-1}\log\Bigg(\left|(1-\rho_1)^\alpha(1-\rho_2)^\alpha - \rho_1^\alpha\rho_2^\alpha\right|^{\frac{1}{\alpha}}\\ &\cdot\left(\left|(1-\rho_2)^\alpha - \rho_1^\alpha\right|^{\frac{1}{1-\alpha}} + \left|(1-\rho_1)^\alpha - \rho_2^\alpha\right|^{\frac{1}{1-\alpha}}\right)^{\frac{\alpha-1}{\alpha}}\Bigg) \end{aligned} \quad (34)$$

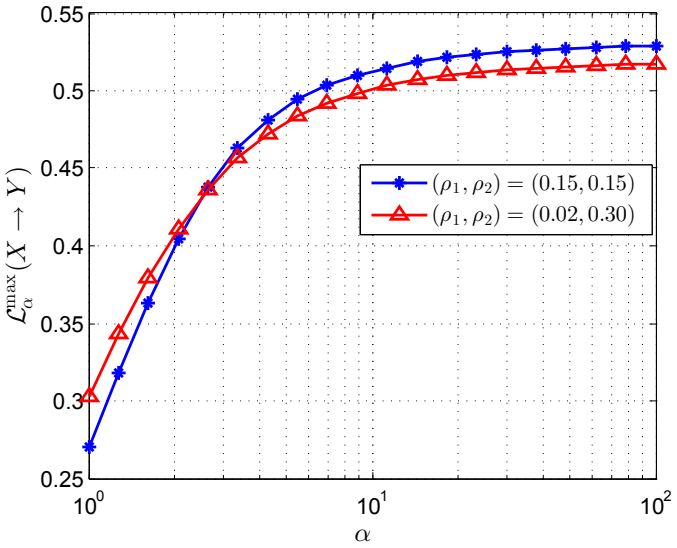


Fig. 4: The values of maximal α-leakage for binary channels determined by a pair of crossover probabilities (ρ_1, ρ_2).

If $\rho_1 = \rho_2$, (34) simplifies to

$$\mathcal{L}_\alpha^{max}(X \to Y) = \frac{1}{\alpha - 1} \log\left((1-\rho_1)^\alpha + \rho_1^\alpha\right) + \log 2, \quad (35)$$

which is exactly the α-leakage for the binary symmetric channel with the uniform input distribution. Fig. 4 plots the values of maximal α-leakage for example channels where $\rho_1 = \rho_2$ and $\rho_1 \neq \rho_2$, and shows that the ordering of leakages for the two channels varies with α.

C. Leakage Measures Based on f-Divergence

We introduce two classes of information leakages derived from f-divergence, called *f-leakage* and *maximal f-leakage*. The f-leakage depends on the distribution of original data, and in contrast, maximal f-divergence only depends on the support of original data. We also show the relation between the f-divergence-based measures and maximal α-leakage for $\alpha = 1$ and $\alpha > 1$, respectively.

Recall that for a convex function $f : \mathbb{R}_+ \to \mathbb{R}$ such that $f(1) = 0$, an f-divergence D_f is a measure of the distance between two distributions given by

$$D_f(P_Y \| Q_Y) = \sum_y Q(y)\, f\left(\frac{P(y)}{Q(y)}\right). \quad (36)$$

Definition 7. *Given a joint distribution $P_{X,Y} = P_{Y|X} P_X$ and a f-divergence D_f, the f-leakage is defined as*

$$\mathcal{L}_f(X \to Y) = \inf_{Q_Y} D_f(P_{X,Y} \| P_X \times Q_Y), \quad (37)$$

and the maximal f-leakage is defined as

$$\mathcal{L}_f^{max}(X \to Y) = \sup_{P_{\tilde{X}}} \inf_{Q_Y} D_f(P_{Y|X} P_{\tilde{X}} \| P_{\tilde{X}} \times Q_Y), \quad (38)$$

where $P_{\tilde{X}}$ is a distribution over the support of P_X.

Note that in Definition 7, maximal f-leakage ($\mathcal{L}_f^{\max}$) depends on the distribution of X only through its support. In contrast, f-leakage ($\mathcal{L}_f$) depends fully on the distribution of X. Both measures depend on the chosen mechanism $P_{Y|X}$.

Recall that for $\alpha = 1$, maximal α-leakage is MI. Therefore, it is a special case of $\mathcal{L}_f(X \to Y)$ in (37) with $f(t) = t \log t$. Furthermore, for $\alpha > 1$, maximal α-leakage has a one-to-one relationship with a special case of $\mathcal{L}_f^{\max}$ in (38) for f given by

$$f_\alpha(t) = \frac{1}{\alpha - 1}(t^\alpha - 1), \quad (39)$$

such that D_f is the Hellinger divergence of order α [51]. The following lemma makes this observation precise.

Lemma 2. *For discrete random variables X and Y, the maximal α-leakage ($\alpha > 1$) from X to Y can be written as*

$$\mathcal{L}_\alpha^{max}(X \to Y) = \tfrac{1}{\alpha - 1} \log\left(1 + (\alpha - 1)\mathcal{L}_{f_\alpha}^{max}(X \to Y)\right), \quad (40)$$

where $\mathcal{L}_{f_\alpha}^{max}(X \to Y)$ indicates a set of maximal f-leakage in (38) defined from the function given by (39).

A detailed proof is in Appendix D. Note that substituting f_α defined in (39) into (36), we can obtain the Hellinger divergences of order $\alpha > 1$. Thus, from Lemma 2, for $\alpha > 1$, maximal α-leakage can be transformed to the maximal f-leakage based on Hellinger divergences via a one-to-one mapping. In this sense, maximal α-leakage is a special case of maximal f-leakage.

IV. Properties of Maximal α-Leakage

Thm. 1 shows that α-leakage is exactly Arimoto MI, and therefore, several basic properties of α-leakage have been shown including (i) non-negativity [43, Sec. II-A], (ii) quasi-convexity[3] in $P_{Y|X}$ given P_X [52, Chapter 3.5], and (iii) post-processing inequality[4] [53, Cor. 1]. We now explore proprieties of maximal α-leakage and show that its properties include: (i) quasi-convexity in the conditional distribution $P_{Y|X}$; (ii) data processing inequalities; (iii) sub-additivity (composition property [28]) and additivity for memoryless mechanisms.

The following theorem results from the expression of maximal α-leakage in Thm. 2 as well as some known properties of Sibson MI [43], [45], [48].

Theorem 3. *For $1 \leq \alpha \leq \infty$, maximal α-leakage*

1. *is quasi-convex in $P_{Y|X}$;*
2. *is monotonically non-decreasing in α;*
3. *satisfies data processing inequalities: let random variables X, Y, Z form a Markov chain, i.e., $X - Y - Z$, then*
$$\mathcal{L}_\alpha^{max}(X \to Z) \leq \mathcal{L}_\alpha^{max}(X \to Y) \quad (41a)$$
$$\mathcal{L}_\alpha^{max}(X \to Z) \leq \mathcal{L}_\alpha^{max}(Y \to Z). \quad (41b)$$
4. *satisfies*
$$\mathcal{L}_\alpha^{max}(X \to Y) \geq 0 \quad (42)$$

[3]For $\alpha \geq 1$ and P_X, the Arimoto MI $I_\alpha^{\mathrm{A}}(X;Y)$ is the logarithm of a linear combination of the p-norm ($p = \alpha$) $\|P_{Y|X}(\cdot|x)\|_\alpha$. From [52, Chapter 3.5], we know a log-convex function is quasi-convex such that $I_\alpha^{\mathrm{A}}(X;Y)$ is quasi-convex in $P_{Y|X}$ given P_X.

[4]From the monotonicity of conditional Arimoto entropy [53, Cor. 1], one can derive that for a Markov chain $X - Y - Z$, $I_\alpha^{\mathrm{A}}(X;Z) \leq I_\alpha^{\mathrm{A}}(X;Y)$.

with equality if and only if X is independent of Y, and

$$\mathcal{L}_\alpha^{max}(X \to Y) \leq \begin{cases} \log|\mathcal{X}| & \alpha > 1 \\ H(P_X) & \alpha = 1 \end{cases} \tag{43}$$

with equality if and only if X is a deterministic function of Y.

A detailed proof is in Appendix E.

Remark 1. *Note that:*

- *Since both MI and MaxL are convex in $P_{Y|X}$, $\mathcal{L}_1^{max}(X \to Y)$ and $\mathcal{L}_\infty^{max}(X \to Y)$ are convex in $P_{Y|X}$.*
- *From the monotonicity in Part 2, we can bound maximal α-leakage from above by*[5]

$$\mathcal{L}_\alpha^{max}(X \to Y) \leq \mathcal{L}_{MaxL}(X \to Y) = I_\infty^S(X;Y). \tag{44}$$

- *The data processing inequalities in (41a) and (41b) are called post-processing inequality and linkage inequality, respectively [54], [55]. It is worth noting that not all information leakage measures satisfy the linkage inequality [25], [55]. Examples include α-leakage, maximal information leakage [18], probability of correctly guessing, and DP.*
- *From the monotonicity of maximal α-leakage and the upper bound of MaxL in [28, Lemma 1], we know that if $|\mathcal{Y}| < |\mathcal{X}|$, the upper bound in (43) can be tighter as*

$$\mathcal{L}_\alpha^{max}(X \to Y) \leq \begin{cases} \log|\mathcal{Y}| & \alpha > 1 \\ \min\{H(P_X), \log|\mathcal{Y}|\} & \alpha = 1, \end{cases}$$

with equality for $\alpha > 1$ if and only if Y is a deterministic function of X.

From Thm. 2, we know that for $\alpha > 1$, maximal α-leakage is the supremum of Arimoto/Sibson MI over all possible distributions on the support of original data, and therefore, is a function of a conditional probability distribution. The following theorem bounds the supremum from below by a closed-form expression of the conditional probability distribution.

Theorem 4 (Lower Bound)**.** *For $1 < \alpha \leq \infty$, maximal α-leakage is bounded from below by*

$$\mathcal{L}_\alpha^{max}(X \to Y) \geq \frac{\alpha}{\alpha - 1} \log \frac{\sum_{y \in \mathcal{Y}} \|P_{Y|X}(y|\cdot)\|_\alpha}{|\mathcal{X}|^{\frac{1}{\alpha}}}, \tag{45}$$

with equality if and only if for all $x_1, x_2 \in \mathcal{X}$, there is

$$\sum_y \frac{P_{Y|X}(y|x_1)^\alpha}{\|P_{Y|X}(y|\cdot)\|_\alpha^{\alpha-1}} = \sum_y \frac{P_{Y|X}(y|x_2)^\alpha}{\|P_{Y|X}(y|\cdot)\|_\alpha^{\alpha-1}}. \tag{46}$$

A detailed proof is in Appendix F.

When data may be revealed multiple times (e.g., entering a password multiple times), it is essential to quantify how mechanisms are designed with maximal α leakage compose in terms of total leakage. Consider two released versions Y_1 and Y_2 of X. The following theorem limits maximal α-leakage to an adversary who has access to both Y_1 and Y_2 simultaneously.

Theorem 5 (Sub-additivity/Composition)**.** *Given a Markov chain $Y_1 - X - Y_2$, we have ($\alpha \in [1, \infty]$)*

$$\mathcal{L}_\alpha^{max}(X \to Y_1, Y_2) \leq \sum_{i \in \{1,2\}} \mathcal{L}_\alpha^{max}(X \to Y_i). \tag{47}$$

A detailed proof is in Appendix G.

The following theorem shows the additivity of maximal α-leakage for memoryless mechanisms.

Theorem 6 (Additivity for Memoryless Mechanisms)**.** *For $\alpha \in [1, \infty]$ and a finite integer $n > 0$, let X^n and Y^n be n-length input and output, respectively, of a memoryless mechanism with no feedback, i.e.,*

$$P_{Y^n|X^n} = \prod_{i=1}^n P_{Y_i|X_i}, \tag{48}$$

where X_i and Y_i represent the i^{th} element of X^n and Y^n, respectively, such that

(1) *for $\alpha > 1$*

$$\mathcal{L}_\alpha^{max}(X^n \to Y^n) = \sum_{i=1}^n \mathcal{L}_\alpha^{max}(X_i \to Y_i) \tag{49}$$

(2) *for $\alpha = 1$*

$$\mathcal{L}_1^{max}(X^n \to Y^n) \leq \sum_{i=1}^n \mathcal{L}_1^{max}(X_i \to Y_i) \tag{50}$$

with equality if and only if entries of X^n are mutually independent.

A detailed proof is in Appendix H.

V. Privacy-Utility Tradeoff with a Hard Distortion Constraint

In a privacy-guaranteed data publishing setting, a data curator/provider uses a mapping called *privacy mechanism* to generate distorted versions of original data for releases. The privacy mechanism determines the fidelity of the released data. With a higher fidelity, more utility is maintained, while less privacy preserved. Therefore, a privacy-utility tradeoff (PUT) problem arises in the design of the privacy mechanism.

We consider the two different data publishing scenarios shown in Figs. 1a and 1b: the first where the entirety of the dataset X is considered private, and the second where the dataset consists of two parts S and X, where only S is considered private. For the first case (Fig. 1a), we use maximal α-leakage as the privacy measure, thereby limiting the inference of any private information about the dataset represented by the function U. For the second case (Fig. 1b), we use α-leakage as the privacy measure, thereby limited the inference only of the specific private information represented by S.

We measure utility in terms of a *hard distortion* measure, which constrains the privacy mechanism so that the distortion between each pair of original and released data is bounded with probability 1. Unlike average distortion measures, the

[5]For $\alpha = \infty$, the $I_\infty^S(P_X, P_{Y|X})$ depends on the marginal distribution P_X only through the support of X.

hard distortion measure gives all data samples the same fidelity guarantee (distortion bound), which is independent of the probabilities of the samples. Therefore, the hard distortion measure excludes the case that large distortions are applied to samples with very small probabilities, which is possible under average distortion constraints. The fidelity guarantee can lead to better performance for applications for which low probability events cannot be ignored or excluded easily (e.g., anomaly detection from released datasets or high-fidelity empirical distribution estimation for census applications) but is incompatible with some privacy measures like DP and L-DP. Specifically, for the original and released data X, Y and a distortion function $d(\cdot,\cdot)$, the utility guarantee is modeled as the hard distortion constraint $d(X,Y) \leq D$ *with probability* 1, where D is the maximal permitted distortion. In other words, if a privacy mechanism $P_{Y|X}$ satisfies the hard distortion constraint, given input x, the output y of the privacy mechanism must lie in a *non-empty* set $B_D(x)$ given by

$$B_D(x) \triangleq \{y : d(x,y) \leq D\}, \tag{51}$$

i.e., for any x with $P_X(x) > 0$, $P_{Y|X}(y|x) = 0$ if $y \notin B_D(x)$. Thus, a mathematical model of the PUT problem is given by

$$\inf_{P_{Y|X} \in \mathcal{P}_{Y|X}} \quad \mathcal{L}^{(\cdot)}_{(\cdot)}(X \to Y) \tag{52a}$$

$$\text{s.t.,} \quad d(X,Y) \leq D, \tag{52b}$$

where the set $\mathcal{P}_{Y|X}$ is the collection of stochastic matrices, and the superscript and subscript of $\mathcal{L}$ depend on the privacy measure under consideration (see Sec. III for notation).

Remark 2. *Note that given any input x, the hard distortion constraint in (52b) will force the conditional probabilities of the outputs that are not in $B_D(x)$ to be zero. Thus, this utility guarantee is* incompatible *with some privacy notions, which require each input to be mapped to all outputs with some positive probabilities; e.g., DP and any maximal f-leakage with $f(0) = \infty$.*

A. PUTs for Entirely Sensitive Datasets

For the privacy-guaranteed publishing of an entirely sensitive dataset shown in Fig. 1a, we use maximal α-leakage as the privacy measure. From Section III-C, we know that maximal α-leakage is a specific case of f-leakage and maximal f-leakage (in Def. 7) for $\alpha = 1$ and $\alpha > 1$, respectively. Hereby, we solve the PUT problems which minimize either f-leakage or maximal f-leakage, subject to a hard distortion constraint. By applying the relations between maximal α-leakage and the f-divergence-based variants, we derive the optimal PUTs and optimal privacy mechanisms for the PUT problem with maximal α-leakage as the privacy measure. We denote an optimal PUT as $\text{PUT}_{\text{HD},\mathcal{L}^{(\cdot)}_{(\cdot)}}$, where HD and $\mathcal{L}^{(\cdot)}_{(\cdot)}$ in the subscript indicate the hard distortion and the involved privacy measure, respectively.

The following theorem characterizes the optimal tradeoff, i.e., the minimal leakage for any given distortion bound D, denoted as $\text{PUT}_{\text{HD},\mathcal{L}_f}(D)$, in (52) for the case that f-leakage is used as the privacy measure.

Theorem 7. *For any f-leakage $\mathcal{L}_f$ in (37) and a distortion function $d(\cdot,\cdot)$ with $B_D(x)$ in (51), the optimal PUT in (52) is given by*

$$PUT_{HD,\mathcal{L}_f}(D)$$

$$= \inf_{P_{Y|X}: d(X,Y) \leq D} \mathcal{L}_f(X;Y), \tag{53}$$

$$= f(0) + \inf_{Q_Y} \mathbb{E}\Big[Q_Y(B_D(X))\Big(f\big(\tfrac{1}{Q_Y(B_D(X))}\big) - f(0)\Big)\Big]. \tag{54}$$

Moreover, letting $Q_Y^\star$ be the distribution achieving the infimum in (54), an optimal mechanism $P^\star_{Y|X}$ is given by

$$P^\star_{Y|X}(y|x) = \frac{\mathbf{1}\big(d(x,y) \leq D\big) Q^\star_Y(y)}{Q^\star_Y(B_D(x))}. \tag{55}$$

A detailed proof in Appendix I. Note that as a result of the distribution dependence of the leakage measure $\mathcal{L}_f$ in (37), the optimal tradeoff in (54) is an *expected* function of X. The optimal mechanism $P^*_{Y|X}(y|x)$ is, in fact, the normalized probability of y when the conditional support of Y given $X = x$ is restricted to $B_D(x)$ (i.e., Y is restricted to taking values in $B_D(x)$ for a given x).

In (52), making use of maximal f-divergence as the privacy measure, the optimal PUT, denoted as $\text{PUT}_{\text{HD},\mathcal{L}_f^{\max}}(D)$, with respect to the hard distortion constraint is given as the minimal leakage for any given distortion bound D in the following theorem.

Theorem 8. *For any maximal f-leakage $\mathcal{L}_f^{max}$ in (38), a distortion function $d(\cdot,\cdot)$ and $B_D(x)$ in (51), the optimal PUT in (52) is given by*

$$PUT_{HD,\mathcal{L}_f^{max}}(D) = \inf_{P_{Y|X}: d(X,Y) \leq D} \mathcal{L}_f^{max}(X \to Y), \tag{56}$$

$$= q^\star f((q^\star)^{-1}) + (1 - q^\star) f(0), \tag{57}$$

with $q^\star$ defined as

$$q^\star \triangleq \sup_{Q_Y} \inf_x Q_Y(B_D(x)). \tag{58}$$

Moreover, letting $Q_Y^\star$ be the distribution achieving the supremum in (58), an optimal mechanism $P^\star_{Y|X}$ is given by (55).

A detailed proof is in Appendix J. Observe that, in contrast to the optimal tradeoff $\text{PUT}_{\text{HD},\mathcal{L}_f}$ for f-leakage in Thm. 7, which depends on the probability distribution P_X, the optimal tradeoff $\text{PUT}_{\text{HD},\mathcal{L}_f^{\max}}$ for maximal f-leakage depends only on the support of P_X. This results from the fact that the maximal f-leakage $\mathcal{L}_f^{\max}$ in (38) is the supremum over all possible probability distribution on the support of P_X, and therefore, depends on P_X only through the support.

The next corollary characterizes the optimal tradeoff $\text{PUT}_{\text{HD},\mathcal{L}_\alpha^{\max}}$ for maximal α-leakage. Recall that for $\alpha = 1$, $\mathcal{L}_1^{\max}$ equals $\mathcal{L}_f$ with $f(t) = t \log t$. For $\alpha > 1$, from the one-to-one relationship between $\mathcal{L}_\alpha^{\max}$ and $\mathcal{L}_{f_\alpha}^{\max}$ in (40), we know that finding $\text{PUT}_{\text{HD},\mathcal{L}_\alpha^{\max}}$ is equivalent to finding the optimal tradeoff $\text{PUT}_{\text{HD},\mathcal{L}_f^{\max}}$ in (56) for $\mathcal{L}_f^{\max} = \mathcal{L}_{f_\alpha}^{\max}$.

Corollary 1. *For maximal α-leakage, the optimal PUT in (52)*

is given by

$$PUT_{HD,\mathcal{L}_\alpha^{max}}(D) = \inf_{P_{Y|X}:d(X,Y)\le D} \mathcal{L}_\alpha^{max}(X\to Y), \tag{59}$$

$$= \begin{cases} \inf_{Q_Y} \mathbb{E}\left[\log \frac{1}{Q_Y(B_D(X))}\right], & \alpha = 1 \quad (60a) \\ -\log q^\star, & \alpha > 1 \quad (60b) \end{cases}$$

where $q^\star$ is defined in (58). Moreover, an optimal mechanism is given by (55), where for $\alpha = 1$, $Q_Y^\star$ achieves the infimum in (60a); and for $\alpha > 1$, $Q_Y^\star$ achieves the supremum in (58).

Remark 3. *The optimal PUTs in (54) and (57) simplify to finding an output distribution Q_Y^* that can be viewed as a "target" distribution, i.e., the optimal mechanism aims to produce this distribution as closely as possible, subject to the utility constraint. In particular, given an input, the optimal mechanism in (55) distributes the outputs according to Q_Y^* while conditioning the output to be within a ball of radius D around the input. The optimization in (58) ensures that all inputs are uniformly masked while (54) provides average guarantees.*

Moreover, for any arbitrarily chosen maximal f-leakage, the optimal PUT in (57) leads to the same target distribution Q_Y^* given by (58). Therefore, the corresponding optimal mechanism in (55) is independent of the choice of maximal f-leakage. As a special case of (57), the optimal tradeoff in (60b), for maximal α-leakage with $\alpha > 1$, is achieved by the optimal mechanism (in (55)) that is no longer depending on α. And the optimal tradeoff (in (60b)) itself is also independent of the value of $\alpha > 1$.

B. PUTs for Datasets Containing Non-Sensitive Data

For datasets containing both sensitive and non-sensitive data, indicated by S and X, respectively, as shown in Fig 1b, the purpose of privacy protection is to limit information leakage of sensitive data while releasing non-sensitive data. We use α-leakage from S to Y as the privacy measure, where Y is the released version of X. Therefore, with $P_{Y|S,X}$ in the place of $P_{Y|X}$ in (52), we obtain the optimal PUT as

$$\text{PUT}_{\text{HD},\mathcal{L}_\alpha}(D) = \inf_{P_{Y|S,X}:d(X,Y)\le D} \mathcal{L}_\alpha(S;Y). \tag{61}$$

The following theorem bounds $\text{PUT}_{\text{HD},\mathcal{L}_\alpha}$ from below. Note that the B_D in the following is the distortion ball defined in (51).

Theorem 9. *The minimal leakage $PUT_{HD,\mathcal{L}_\alpha}$ $(1 \le \alpha \le \infty)$ in (61) is bounded from below by*

$$PUT_{HD,\mathcal{L}_\alpha}(D) \ge$$

$$\begin{cases} \sum_{s,x} P(s,x) \log \left(\max_{y\in B_D(x)} \sum_{s'\in\mathcal{S}_D(y)} P(s') \right)^{-1}, & \alpha = 1 \\ \log \sum_{s,x} \frac{P(s)P(s,x)}{\max_s P_S(s)} \left(\max_{y\in B_D(x)} \sum_{s'\in\mathcal{S}_D(y)} P(s') \right)^{-1}, & \alpha = \infty \\ \frac{\alpha}{\alpha-1} \log \sum_{s,x} \frac{P(s)^\alpha P(x|s)}{\|P_S\|_\alpha} \left(\max_{y\in B_D(x)} \sum_{s'\in\mathcal{S}_D(y)} P(s')^\alpha \right)^{\frac{1-\alpha}{\alpha}}, & \text{else} \end{cases}$$

where the set $\mathcal{S}_D(y)$ of s for each y is defined as

$$\mathcal{S}_D(y) \triangleq \{s : \exists\, x,\ P_{S,X}(s,x) > 0, d(x,y) \le D\}. \tag{62}$$

The lower bound is tight if there exists an privacy mechanism $P_{Y|S,X} \in \mathcal{P}_{Y|S,X}(D)$ such that

(i) *given (s,x), for any y with $P(y|s,x) > 0$,*

$$\sum_{s'\in\mathcal{S}_D(y)} P(s') = \max_{y'\in B_D(s,x)} \sum_{s'\in\mathcal{S}_D(y')} P(s'); \tag{63}$$

(ii) *given any y with $P_Y(y) > 0$, for any $s \in S_D(y)$,*

$$\sum_{x:d(x,y)\le D} P(y|s,x)P(x|s) = \frac{P_Y(y)}{\sum_{s'\in\mathcal{S}_D(y)} P(s')}, \tag{64}$$

where P_Y is the marginal distribution of Y from the privacy mechanism $P_{Y|S,X}$ and $P_{S,X}$.

The proof details are in Appendix K.

Note that by using maximal α-leakage as the privacy measure, the setting for publishing datasets consisting of sensitive and non-sensitive data can be generalized to restrict leakages about *all* functions of the sensitive data. This will be addressed in future work.

VI. Applications: PUTs for Hard and Average Distortion Constraints

In this section, we first illustrate the results of Sec. V and present the optimal PUTs for two distinct hard distortion functions. Our first choice for hard distortion, restricted to binary datasets, is the absolute distance between the types (i.e., empirical distributions) of the original and revealed (binary) datasets. This choice is motivated by the observation that, for any dataset, the type is a sufficient statistic for any function of the dataset that is unaffected by permutation—for example, mean, variance, correlation between two features. Thus, constraining the distortion between the released type and the original type, one can guarantee the utility of the released dataset for a variety of statistical applications. In Example 1 below, we derive the optimal mechanism under this distortion measure for binary datasets. Our second choice for hard distortion is the Hamming distance between the original and released datasets for discrete alphabets. This choice is motivated by the fact that a hard Hamming distortion is more relevant when the order of the entries in the dataset cannot be changed. In Example 2 below, we derive the optimal mechanism under this distortion for a dataset sampled from an arbitrary discrete alphabet. Note that, as a consequence of Corollary 1, for these examples the optimal PUT and privacy mechanism are the same for all values of $\alpha > 1$.

In contrast to hard distortion measures, in Example 3, we study the PUTs that result from using *average* Hamming distortion as the utility measure and maximal α-leakage as the privacy measure. Due to lack of closed-form solutions, we use numerical results to highlight the dependence of both the optimal PUTs and the privacy mechanisms on α.

A. Example 1: Binary Datasets with Hard Distortion on Types

Let X^n be a random dataset with n entries and Y^n be the corresponding released dataset generated by a privacy mechanism $P_{Y^n|X^n}$. Entries of both X^n and Y^n are from the same alphabet $\mathcal{X}$. Adopting the notation of [56, Chapter 11], let P_{x^n} and P_{y^n} indicate the types of input dataset x^n and output dataset y^n, respectively. We define the distortion function as the distance between types, given by

$$d_{\mathrm{T}}(x^n, y^n) = \max_{x\in\mathcal{X}} |P_{x^n}(x) - P_{y^n}(x)|, \tag{65}$$

and therefore, obtain $\mathrm{PUT}_{\mathrm{HD},\mathcal{L}_\alpha^{\max}}$ as in (59) but with datasets X^n, Y^n in place of single letters X, Y. Since types of n-length sequences take on only values that are multiples of $\frac{1}{n}$, this distortion function d_{T} takes on values of the form $\frac{m}{n}$, where $m \in [0, n]$.

We concentrate on binary datasets, i.e., $\mathcal{X} = \{0, 1\}$. Note that for binary datasets, we can simply write $d_{\mathrm{T}}(x^n, y^n) = |P_{x^n}(1) - P_{y^n}(1)|$. For a n-length binary dataset, the number of types is $n+1$. Therefore, all input and output datasets can be categorized into $n+1$ type classes defined as

$$T(i) \triangleq \{x^n : nP_{x^n}(1) = i\}. \tag{66}$$

Theorem 10. *For binary datasets and the distortion function in (65), given integers n, m where $0 \le m \le n$, the optimal tradeoff for $\alpha > 1$ is*

$$PUT_{HD,\mathcal{L}_\alpha^{max}}\left(\frac{m}{n}\right) = \min_{\substack{P_{Y^n|X^n}:\\ d_T(X^n,Y^n)\le\frac{m}{n}}} \mathcal{L}_\alpha^{max}(X^n \to Y^n) \tag{67}$$

$$= \log\left\lceil \frac{n+1}{2m+1} \right\rceil. \tag{68}$$

An optimal privacy mechanism maps all input datasets in a type class to a unique output dataset which is feasible and belongs to a type class in the set $\mathcal{T}^\star$ given by

$$\mathcal{T}^\star \triangleq \left\{ T(j) : j = l + (2m+1)k, k \in \left[0, \left\lceil \tfrac{n+1}{2m+1} \right\rceil - 1\right] \right\}, \tag{69}$$

where $l = m$ if $\lceil \frac{n+1}{2m+1} \rceil - \frac{n+1}{2m+1} \le \frac{m}{2m+1}$, and otherwise, $l = n - \left(\lceil \frac{n+1}{2m+1} \rceil - 1\right)(2m+1)$.

A detailed proof is in Appendix L. Note that for any x^n in the type class $T(i)$, the corresponding output y^n generated by the optimal mechanism $P^*_{Y^n|X^n}$ is unique and belongs to the unique type class in $\mathcal{T}^\star \cap \{T(j) : |i - j| \le m\}$. For example, if $(n, m) = (9, 2)$, then from Thm. 10, we have $\mathrm{PUT}_{\mathrm{HD},\mathcal{L}_\alpha^{\max}}(\frac{2}{9}) = 1$ bit and $\mathcal{T}^\star = \{T(2), T(7)\}$. Fig. 5 shows the optimal mechanism, which maps all input datasets in $\{T(i) : i \in [0, 4]\}$ (resp. $\{T(i) : i \in [5, 9]\}$) to a *unique output dataset* in $T(2)$ (resp. $T(7)$) with probability 1.

B. Example 2: Hard Hamming Distortion on Datasets

In the example, we consider hard Hamming distortion on datasets with entries from general finite alphabets. Formally, for datasets $x^n, y^n \in \mathcal{X}^n$, we define the Hamming distortion function as

$$d_{\mathrm{H}}(x^n, y^n) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(x_i \ne y_i). \tag{70}$$

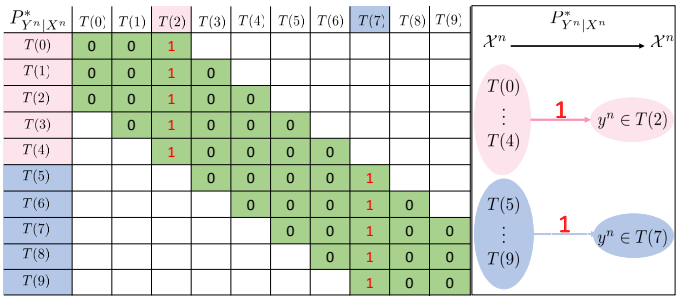


Fig. 5: An optimal mechanism $\mathrm{PUT}_{\mathrm{HD},\mathcal{L}_\alpha^{\max}}\left(\frac{m}{n}\right)$ for $\alpha > 1$ with $(n, m) = (9, 2)$, where rows and columns are types of X^n and Y^n, respectively. Note that the hard distortion forces conditional probabilities of outputs outside the feasible ball of given input to be zero. We highlight the conditional probabilities of feasible outputs in green, and give their values in the optimal mechanism.

Therefore, we obtain $\mathrm{PUT}_{\mathrm{HD},\mathcal{L}_\alpha^{\max}}$ as in (59) but with datasets X^n, Y^n in place of single letters X, Y.

Theorem 11. *For datasets from a finite alphabet $\mathcal{X}$ and Hamming distortion function, for any integers n, m where $0 \le m \le n$, the optimal tradeoff for $\alpha > 1$ is*

$$PUT_{HD,\mathcal{L}_\alpha^{max}}\left(\frac{m}{n}\right) = \min_{\substack{P_{Y^n|X^n}:\\ d_H(x^n,y^n)\le\frac{m}{n}}} \mathcal{L}_\alpha^{max}(X^n \to Y^n) \tag{71}$$

$$= \log \frac{|\mathcal{X}|^n}{\sum_{i=0}^{m}\binom{n}{i}(|\mathcal{X}|-1)^i}. \tag{72}$$

*An optimal privacy mechanism maps each input $x^n \in \mathcal{X}^n$ **uniformly** to every feasible output, i.e., for all x^n, y^n where $d_H(x^n, y^n) \le \frac{m}{n}$, $P_{Y^n|X^n}(y^n|x^n) = \frac{1}{\sum_{i=0}^{m}\binom{n}{i}(|\mathcal{X}|-1)^i}$.*

Note that for any pair of x^n and y^n, the optimal mechanism $P^*_{Y^n|X^n}(y^n|x^n)$ is the average probability of y^n when the support of Y^n is restricted to $B_D(x_n)$, i.e., Y^n takes values from $\{y^n : d_{\mathrm{H}}(x^n, y^n) \le \frac{m}{n}\}$. The key observation to reach the conclusion in Thm. 11 is that every output dataset is in the same number of feasible balls, such that a uniform distribution over the output space leads to equal probability for the feasible ball of each input dataset. The proof details are in Appendix M. Fig. 6 illustrates the optimal mechanism in Thm. 11 for $\mathcal{X} = \{0, 1, 2\}$ and $(n, m) = (2, 1)$.

Note that permuting items of a dataset does not change the type but will lead to a non-zero Hamming distortion. The distortion on types in (65) can be viewed as a relaxation of the Hamming distortion, in the sense that the set of feasible privacy mechanisms in (71) belongs to that in (67), i.e.,

$$\left\{P_{Y^n|X^n} : d_{\mathrm{H}}(x^n, y^n) \le \frac{m}{n}\right\} \subset \left\{P_{Y^n|X^n} : d_{\mathrm{T}}(x^n, y^n) \le \frac{m}{n}\right\}.$$

Therefore, for non-binary alphabets, the result in Thm. 11 limits the minimal leakage in (67).

C. Example 3: Average Hamming Distortion on Binary Alphabet

We consider a PUT setting with maximal α-leakage as the privacy measure and average Hamming distortion as the

$P^*_{Y^n\|X^n}$: X^n \ Y^n	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)	(2,0)	(2,1)	(2,2)
(0,0)	0.2	0.2	0.2	0.2			0.2		
(0,1)	0.2	0.2	0.2		0.2			0.2	
(0,2)	0.2	0.2	0.2			0.2			0.2
(1,0)	0.2			0.2	0.2	0.2	0.2		
(1,1)		0.2		0.2	0.2	0.2		0.2	
(1,2)			0.2	0.2	0.2	0.2			0.2
(2,0)	0.2			0.2			0.2	0.2	0.2
(2,1)		0.2			0.2		0.2	0.2	0.2
(2,2)			0.2			0.2	0.2	0.2	0.2

Fig. 6: An optimal mechanism of (71) for $\alpha > 1$ with $(n, m) = (2, 1)$ and $\mathcal{X} = \{0, 1, 2\}$ where rows and columns are x^n and y^n, respectively. Note that we color the conditional probabilities of feasible outputs (respect to the hard Hamming distortion) and their values are the same as 0.2 in the optimal mechanism.

distortion constraint. Such an average utility constraint can be relevant to data publishing settings where preserving statistics of the dataset is desired. This example also illustrates that, in contrast to the hard distortion constraint, the optimal mechanism may depend on α.

Consider the following PUT problem that minimizes maximal α-leakage subject to the average Hamming distortion constraint:

$$\min_{P_{Y|X}} \quad \mathcal{L}_\alpha^{\max}(X \to Y) \tag{73a}$$

$$\text{s.t.,} \quad \sum_{x,y \in \mathcal{X}} P_{X,Y}(x,y)\mathbf{1}\left(y \neq x\right) \leq D \tag{73b}$$

where $0 < D < 1 - \max_x P_X(x)$ is the maximum permitted average Hamming distortion. We focus on the binary case: let $X, Y \in \{0, 1\}$ where X follow the Bernoulli distribution Bern(p) $(0 < p < 1)$, i.e., $P_X(1) = p$. We represent the privacy mechanism $P_{Y|X}$ via the two crossover probabilities $P_{Y|X}(1|0) = \rho_1$ and $P_{Y|X}(0|1) = \rho_2$. By solving the supremum in the expression of maximal α-leakage, for $\alpha > 1$, the optimization in (73) can be written as

$$\min_{\rho_1,\rho_2} \frac{1}{\alpha - 1} \log\Big((1-\rho_1)^\alpha (1-\rho_2)^\alpha - (\rho_1\rho_2)^\alpha\Big) + \log\Big(\big((1-\rho_1)^\alpha - \rho_2^\alpha\big)^{\frac{1}{1-\alpha}} + \big((1-\rho_2)^\alpha - \rho_1^\alpha\big)^{\frac{1}{1-\alpha}}\Big) \tag{74a}$$

$$\text{s.t.} \quad (1-p)\rho_1 + p\rho_2 \leq D. \tag{74b}$$

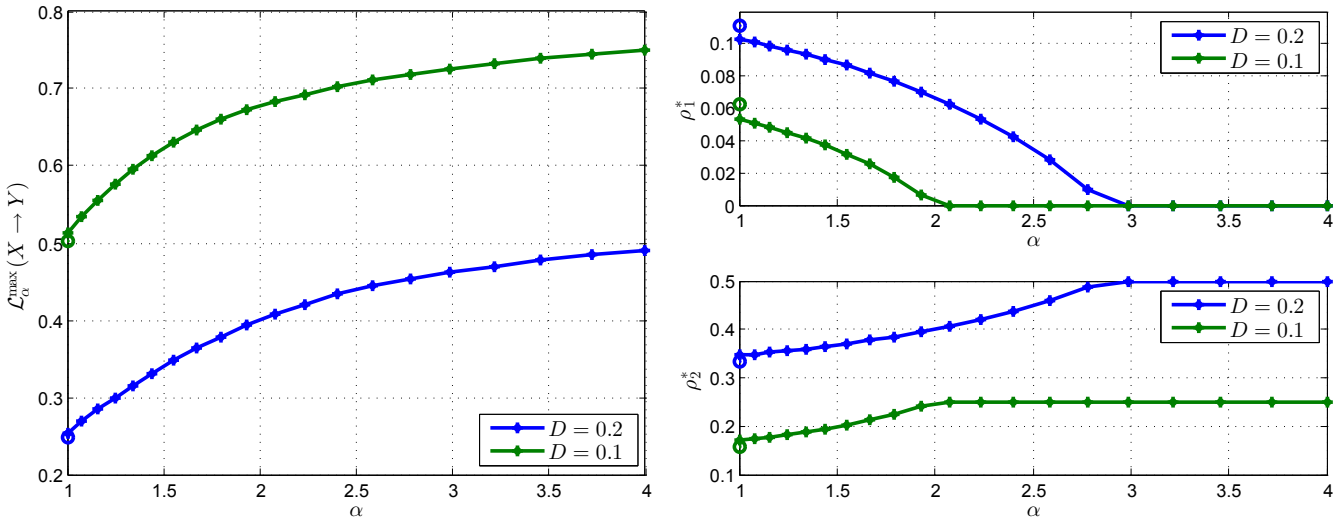
Fig. 7 shows the optimal values and mechanisms in (74) for $p = 0.4$ and $D = 0.2$ or $D = 0.1$. From the plots, we can see that for $\alpha = 1.001$, the optimal mechanism $P^*_{Y|X}$ (represented by ρ_1^* and ρ_2^*) is slightly different from that of mutual information [56, Figure 10.3] due to the fact that as α tends to 1, the limit of maximal α-leakage is Shannon channel capacity instead of mutual information, i.e., $\lim_{\alpha\to 1} \mathcal{L}_\alpha^{\max}(X \to Y) = \lim_{\alpha\to 1} \sup_{P_{\tilde{X}}} I_\alpha^{\mathrm{A}}(\tilde{X}; Y) = \sup_{P_{\tilde{X}}} I(\tilde{X}; Y)$. We also observe that as α grows, the optimal crossover probabilities ρ_1^* and ρ_2^* gradually approach to 0 and $\frac{D}{p}$, respectively. Therefore, for the PUT in (74), maximal α-leakage with different values of $1 < \alpha < \infty$ leads to various optimal privacy mechanisms, which can differ from that for either $\alpha = 1$ or $\alpha = \infty$.

It is not difficult to check that the optimal privacy mechanisms (in Fig. 7b) for different values of α give the same probability of correctly guessing, defined as $\sum_y P_Y(y) \max_x P_{Y|X}(y|x)$ [27], which equals to $1 - D$ in this example. Probability of correctly guessing is, in fact, the average accuracy of estimating the value of original data X from Y when the maximal posterior (MAP) estimator is used. Therefore, if the original data X is released against an adversary who is only interested in the most likely value of X, all values of α will lead to the same privacy guarantee in the sense that the optimal mechanisms give the same average accuracy of estimation.

VII. Conclusion

Via α-loss $(1 \leq \alpha \leq \infty)$, we have defined two tunable measures of information leakage: α-leakage for a specific function of original data, and maximal α-leakage for any arbitrary function of original data, and proven that: (i) α-leakage equals to Arimoto mutual information for $1 \leq \alpha \leq \infty$; (ii) for $\alpha > 1$, maximal α-leakage equals to Arimoto channel capacity; and for $\alpha = 1$ and $\alpha = \infty$ it simplifies to mutual information and maximal leakage, respectively. From properties of Arimoto mutual information, α-leakage is known to be quasi-convex in the conditional distribution and satisfy the post-processing inequality. For maximal α-leakage, we have proven that it is quasi-convex in the conditional distribution, and satisfies data processing inequalities as well as a composition property.

In the context of privacy-guaranteed data publishing, we have explored PUT problems for the proposed tunable leakage measures and hard distortion utility constraints. This utility constraint has the advantage that it allows the data curator/provider to make specific, deterministic guarantees on the quality of the released dataset. For maximal α-leakage, we have shown that: (i) for all $\alpha > 1$, we obtain the same optimal privacy mechanism and optimal PUT, both of which are independent of the distribution of the original data; (ii) for $\alpha = 1$, the optimal mechanism differs and depends on the distribution of the original data. In other words, for this hard distortion measure, maximal α-leakage behaves as either mutual information or maximal leakage. We have also demonstrated that this extremal behavior may not hold when the hard distortion constraint is replaced by an average distortion constraint (e.g., average Hamming distortion) and the source alphabet is binary. Future directions include studying PUT problems with average distortion constraints for non-binary alphabets to further explore the impact of α on the design of privacy mechanisms.



(a) (b)

Fig. 7: Numerical results for the privacy-utility tradeoff in (74) with $p = 0.4$ and $D \in \{0.2, 0.1\}$. Figure 7a plots the minimal values of maximal α-leakage as a function of α (circles indicate $\alpha = 1$ and stars are for $1.001 \leq \alpha \leq 4$). Figure 7b illustrates the behavior of the crossover probabilities ρ_1^* and ρ_2^* for the optimal privacy mechanisms as a function of α.

Appendix A: Proof of Lemma 1

For $1 < \alpha < \infty$, the minimal expected value of the α-loss in Definition 3 can be expressed as

$$\begin{aligned}
&\min_{P_{\hat{X}|Y}} \mathbb{E}\left[\ell_\alpha(X, Y, P_{\hat{X}|Y})\right] \\
=&\min_{P_{\hat{X}|Y}} \frac{\alpha}{\alpha - 1}\left(1 - \sum_{x,y} P_{X,Y}(x,y) P_{\hat{X}|Y}(x|y)^{\frac{\alpha-1}{\alpha}}\right) && (75)\\
=&\frac{\alpha}{\alpha - 1}\left(1 - \max_{P_{\hat{X}|Y}} \sum_{x,y} P_{X,Y}(x,y) P_{\hat{X}|Y}(x|y)^{\frac{\alpha-1}{\alpha}}\right) && (76)\\
=&\frac{\alpha}{\alpha - 1}\left(1 - \sum_y P(y) \max_{P_{\hat{X}|Y=y}} \sum_x P(x|y) P_{\hat{X}|Y}(x|y)^{\frac{\alpha-1}{\alpha}}\right). && (77)
\end{aligned}$$

For each y with $P_Y(y) > 0$, the maximization in (77) can be explicitly written as

$$\begin{aligned}
\max_{P_{\hat{X}|Y=y}} \quad & \sum_{x \in \mathcal{X}} P_{X|Y}(x|y) P_{\hat{X}|Y}(x|y)^{\frac{\alpha-1}{\alpha}} && (78a)\\
\text{s.t.} \quad & \sum_{x \in \mathcal{X}} P_{\hat{X}|Y}(x|y) = 1 && (78b)\\
& P_{\hat{X}|Y}(x|y) \geq 0 \quad \text{for all } x \in \mathcal{X}. && (78c)
\end{aligned}$$

For $1 \leq \alpha \leq \infty$, the exponent $\frac{\alpha-1}{\alpha} \geq 0$ such that the problem in (78) is a convex program. Therefore, by using Karush-Kuhn-Tucker (KKT) conditions [52, Chapter 5.5.3], we obtain the optimal value of (78) as

$$\max_{P_{\hat{X}|Y=y}} \sum_x P_{X|Y}(x|y) P_{\hat{X}|Y}(x|y)^{\frac{\alpha-1}{\alpha}} = \|P_{X|Y}(\cdot|y)\|_\alpha \quad (79)$$

with the optimal solution $P^\star_{\hat{X}|Y}$ as

$$P^\star_{\hat{X}|Y}(x|y) = \frac{P_{X|Y}(x|y)^\alpha}{\sum_{x \in \mathcal{X}} P_{X|Y}(x|y)^\alpha} \quad \text{for all } x \in \mathcal{X}. \quad (80)$$

For $\alpha = 1$, the optimal solution is $P^\star_{\hat{X}|Y} = P_{X|Y}$. For $\alpha = \infty$, we have

$$\begin{aligned}
\lim_{\alpha \to \infty} P^\star_{\hat{X}|Y}(x|y) &= \lim_{\alpha \to \infty} \frac{\left(\frac{P_{X|Y}(x|y)}{\max_x P_{X|Y}(x|y)}\right)^\alpha}{\sum_{x \in \mathcal{X}} \left(\frac{P_{X|Y}(x|y)}{\max_x P_{X|Y}(x|y)}\right)^\alpha} && (81)\\
&= \begin{cases} \frac{1}{k(y)}, & x = \arg\max_x P_{X|Y}(x|y) \\ 0, & \text{otherwise}, \end{cases} && (82)
\end{aligned}$$

where the integer $k(y)$ indicates the cardinality of the set $\{x : x = \arg\max_x P_{X|Y}(x|y)\}$.
Applying the optimal solution $P^\star_{\hat{X}|Y}$ to (77), we have

$$\begin{aligned}
&\min_{P_{\hat{X}|Y}} \mathbb{E}\left[\ell_\alpha(X, Y, P_{\hat{X}|Y})\right] \\
&= \begin{cases} \frac{\alpha}{\alpha-1}\left(1 - \sum_y \|P_{X,Y}(Xy)\|_\alpha\right), & \alpha > 1 \\ \sum_{x,y} P_{X,Y}(x,y) \log \frac{1}{P_{X|Y}(x|y)}, & \alpha = 1 \end{cases}, && (83)\\
&= \begin{cases} \frac{\alpha}{\alpha-1}\left(1 - \exp\left(\frac{1-\alpha}{\alpha} H_\alpha^{\mathrm{A}}(X|Y)\right)\right), & \alpha > 1 \\ H(X|Y), & \alpha = 1 \end{cases}. && (84)
\end{aligned}$$

□

Appendix B: Proof of Theorem 1

The expression (27) can be explicitly written as

$$\begin{aligned}
&\mathcal{L}_\alpha(X \to Y) \\
&= \lim_{\alpha' \to \alpha} \frac{\alpha'}{\alpha' - 1} \log\left(\frac{\max_{P_{\hat{X}|Y}} \sum_{x,y} P_{X,Y}(x,y) P_{\hat{X}|Y}(x|y)^{\frac{\alpha'-1}{\alpha'}}}{\max_{P_{\hat{X}}} \sum_x P_X(x) P_{\hat{X}}(x)^{\frac{\alpha'-1}{\alpha'}}}\right) && (85)
\end{aligned}$$

To simplify the expression in (85), we need to solve the two maximizations in the logarithm. From (29), we know that to solve the maximization in the numerator equals to find the minimal expected α-loss. Making use of the result in Lemma 1, we have that for $\alpha' \in (1,\infty)$,

$$\max_{P_{\hat{X}|Y}} \sum_{x,y} P_{X,Y}(x,y) P_{\hat{X}|Y}(x|y)^{\frac{\alpha'-1}{\alpha'}} = \exp\left(\frac{1-\alpha'}{\alpha'} H^{\mathrm{A}}_{\alpha'}(X|Y)\right). \quad (86)$$

Similarly, by applying KKT conditions to the maximization in the denominator, we have that for $\alpha' \in (1,\infty)$

$$\max_{P_{\hat{X}}} \sum_{x\in\mathcal{X}} P_X(x) P_{\hat{X}}(x)^{\frac{\alpha'-1}{\alpha'}} = \exp\left(\frac{1-\alpha'}{\alpha'} H_{\alpha'}(X)\right). \quad (87)$$

Therefore, we have for $\alpha' \in (1,\infty)$

$$\begin{aligned}
&\mathcal{L}_\alpha(X\to Y)\\
=&\frac{\alpha'}{\alpha'-1}\log\exp\left(\frac{1-\alpha'}{\alpha'}\left(H^{\mathrm{A}}_{\alpha'}(X|Y) - H_{\alpha'}(X)\right)\right) && (88)\\
=&I^{\mathrm{A}}_{\alpha'}(X;Y). && (89)
\end{aligned}$$

From the continuous extensions of Arimoto MI for $\alpha = 1$ and ∞, respectively, we have that for $1 \le \alpha \le \infty$, α-leakage equals Arimoto MI.

□

Appendix C: Proof of Theorem 2

From Thm. 1, we have for $1 \le \alpha \le \infty$,

$$\mathcal{L}^{\max}_\alpha(X\to Y) = \sup_{U-X-Y} I^{\mathrm{A}}_\alpha(U;Y). \quad (90)$$

If $\alpha = 1$, we have

$$\mathcal{L}^{\max}_1(X\to Y) = \sup_{U-X-Y} I(U;Y) \le I(X;Y) \quad (91)$$

where the inequality is from data processing inequalities of MI [56, Thm 2.8.1]. We then prove that the upper bound $I(X;Y)$ in (91) can be achieved. Let U be a function of X satisfying $H(X|U) = 0$. From the condition $H(X|U) = 0$ and the Markov chain $U - X - Y$, we have

$$\begin{aligned}
H(X,Y|U) =& H(X|U) + H(Y|X,U) = 0 + H(Y|X),\\
H(X,Y|U) =& H(Y|U) + H(X|Y,U) = H(Y|U) + 0,
\end{aligned}$$

i.e., $H(Y|X) = H(Y|U)$. Therefore, for a function U satisfying $H(X|U) = 0$, there is $\mathcal{L}_1(U\to Y) = I(U;Y) = I(X;Y)$.

If $\alpha = \infty$, we have

$$\mathcal{L}^{\max}_\infty(X\to Y) = \sup_{U-X-Y} \log \frac{\sum_y P_Y(y) \max_u P_{U|Y}(u|y)}{\max_u P_U(u)}, \quad (92)$$

which is exactly the expression of MaxL, and therefore, we have that for $\alpha = \infty$, the maximal α-leakage equals to the Sibson MI of order ∞ [28, Thm. 1], i.e.,

$$\mathcal{L}^{\max}_\infty(X\to Y) = \log \sum_y \max_x P_{Y|X}(y|x). \quad (93)$$

For $\alpha \in (1,\infty)$, we provide an upper bound for $\mathcal{L}^{\max}_\alpha(X\to Y)$, and then, give an achievable scheme as follows.

Upper Bound: We have an upper bound of $\mathcal{L}^{\max}_\alpha(X\to Y)$ as

$$\begin{aligned}
\mathcal{L}^{\max}_\alpha(X\to Y) =& \sup_{U-X-Y} I^{\mathrm{A}}_\alpha(U;Y) && (94)\\
=& \sup_{\substack{P_{Y,\tilde{X}|\tilde{U}}: P_{\tilde{X}} = P_X\\ P_{Y|\tilde{X},\tilde{U}} = P_{Y|X}}} \sup_{P_{\tilde{U}}} I^{\mathrm{A}}_\alpha(\tilde{U};Y) && (95)\\
\le& \sup_{\substack{P_{Y,\tilde{X}|\tilde{U}}: P_{\tilde{X}|\tilde{U}}(\cdot|u) \ll P_X\\ P_{Y|\tilde{X},\tilde{U}} = P_{Y|X}}} \sup_{P_{\tilde{U}}} I^{\mathrm{A}}_\alpha(\tilde{U};Y) && (96)\\
=& \sup_{\substack{P_{Y,\tilde{X}|\tilde{U}}: P_{\tilde{X}|\tilde{U}}(\cdot|u) \ll P_X\\ P_{Y|\tilde{X},\tilde{U}} = P_{Y|X}}} \sup_{P_{\tilde{U}}} I^{\mathrm{S}}_\alpha(\tilde{U};Y) && (97)\\
=& \sup_{P_{\tilde{X}} \ll P_X} I^{\mathrm{S}}_\alpha(\tilde{X};Y) && (98)\\
=& \sup_{P_{\tilde{X}} \ll P_X} I^{\mathrm{A}}_\alpha(\tilde{X};Y) && (99)
\end{aligned}$$

where $P_{\tilde{X}} \ll P_X$ indicate that the support of $P_{\tilde{X}}$ is a subset of the support of P_X[6]. In (95), $\tilde{U} - \tilde{X} - Y$ forms a Markov chain and the probability distribution of $\tilde{X}$ is constrained to be P_X. The upper bound in (96) results from allowing $\tilde{X}$ to be distributed arbitrarily over the support of X. The equations in (97) and (99) result from that Arimoto MI and Sibson MI of order $\alpha > 0$ have the same supremum [43, Thm. 5], which can be proved from the expressions of Arimoto and Sibson MIs as follows:

$$\begin{aligned}
&\sup_{P_{\tilde{U}}} I^{\mathrm{A}}_\alpha(\tilde{U};Y)\\
=&\sup_{P_{\tilde{U}}} \frac{\alpha}{\alpha-1}\log \frac{\sum_y \left(\sum_u P_{\tilde{U},Y}(u,y)^\alpha\right)^{\frac{1}{\alpha}}}{\left(\sum_u P_{\tilde{U}}(u)^\alpha\right)^{\frac{1}{\alpha}}} && (100)\\
=&\sup_{P_{\tilde{U}}} \frac{\alpha}{\alpha-1}\log \sum_y \left(\sum_u \frac{P_{\tilde{U}}(u)^\alpha}{\sum_u P_{\tilde{U}}(u)^\alpha} P_{Y|\tilde{U}}(y|u)^\alpha\right)^{\frac{1}{\alpha}} && (101)\\
=&\sup_{P_{\tilde{U}'}} \frac{\alpha}{\alpha-1}\log \sum_y \left(\sum_u P_{\tilde{U}'}(u) P_{Y|\tilde{U}}(y|u)^\alpha\right)^{\frac{1}{\alpha}} && (102)\\
=&\sup_{P_{\tilde{U}'}} I^{\mathrm{A}}_\alpha(\tilde{U}';Y) && (103)
\end{aligned}$$

where $P_{\tilde{U}}$ and $P_{\tilde{U}'}$ are probability distributions over the same support and for each u, $P_{\tilde{U}'}(u) = \frac{P_{\tilde{U}}(u)^\alpha}{\sum_u P_{\tilde{U}}(u)^\alpha}$. From the data processing inequalities of Sibson MI for the Markov chain $\tilde{U} - \tilde{X} - Y$, we have that $I^{\mathrm{S}}_\alpha(\tilde{U};Y) \le I^{\mathrm{S}}_\alpha(\tilde{X};Y)$ with equality if and only if $\tilde{U} = \tilde{X}$ [43, Thm. 3]. Therefore, in (97) $\sup_{P_{\tilde{U}}} I^{\mathrm{S}}_\alpha(\tilde{U};Y) = I^{\mathrm{S}}_\alpha(\tilde{X};Y)$, and then, by replacing $\tilde{U}$ with $\tilde{X}$ we have (98).

Lower bound: We bound (90) from below by considering a random variable U such that $U - X - Y$ is a Markov chain and $H(X|U) = 0$. Specifically, let the alphabet $\mathcal{U}$ consist of $\mathcal{U}_x$, a collection of U mapped to a $x \in \mathcal{X}$, i.e., $\mathcal{U} = \bigcup_{x\in\mathcal{X}} \mathcal{U}_x$

[6]Note that any set is also the subset of itself, such that the support of $\tilde{X}$ can be the same as that of X

with $U = u \in \mathcal{U}_x$ if and only if $X = x$. Therefore, for the specific variable U, we have

$$P_{Y|U}(y|u) = \begin{cases} P_{Y|X}(y|x) & \text{for all } u \in \mathcal{U}_x \\ 0 & \text{otherwise.} \end{cases} \tag{104}$$

Construct a probability distribution $P_{\tilde{X}}$ over $\mathcal{X}$ from P_U as

$$P_{\tilde{X}}(x) = \frac{\sum_{u\in\mathcal{U}_x} P_U(u)^\alpha}{\sum_{x\in\mathcal{X}}\sum_{u\in\mathcal{U}_x} P_U(u)^\alpha} \quad \text{for all } x \in \mathcal{X}. \tag{105}$$

Thus,

$$\begin{aligned}
&I_\alpha^{\mathrm{A}}(U;Y) \\
&= \frac{\alpha}{\alpha-1}\log \frac{\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}}\sum_{u\in\mathcal{U}_x} P_{Y|U}(y|u)^\alpha P_U(u)^\alpha\right)^{\frac{1}{\alpha}}}{\left(\sum_{x\in\mathcal{X}}\sum_{u\in\mathcal{U}_x} P_U(u)^\alpha\right)^{\frac{1}{\alpha}}} && (106)\\
&= \frac{\alpha}{\alpha-1}\log \sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P_{Y|X}(y|x)^\alpha \frac{\sum_{u\in\mathcal{U}_x} P_U(u)^\alpha}{\sum_{x\in\mathcal{X}}\sum_{u\in\mathcal{U}_x} P_U(u)^\alpha}\right)^{\frac{1}{\alpha}} && (107)\\
&= \frac{\alpha}{\alpha-1}\log \sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P_{Y|X}(y|x)^\alpha P_{\tilde{X}}(x)\right)^{\frac{1}{\alpha}} && (108)\\
&= I_\alpha^{\mathrm{S}}(\tilde{X};Y). && (109)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathcal{L}_\alpha^{\max}(X\to Y) &= \sup_{U-X-Y} I_\alpha^{\mathrm{A}}(U;Y) && (110)\\
&\geq \sup_{U:U-X-Y, H(X|U)=0} I_\alpha^{\mathrm{A}}(U;Y) && (111)\\
&= \sup_{P_{\tilde{X}}\ll P_X} I_\alpha^{\mathrm{S}}(\tilde{X};Y), && (112)
\end{aligned}$$

where the last equality follows because, for any $P_{\tilde{X}} \ll P_X$, there exists a distribution $P_U(u)$ for $u \in \mathcal{U}$ such that (105) holds; therefore, the supremum over these U in (104) is equivalent to the supremum of $P_{\tilde{X}}$. Therefore, combining (98) and (112), we obtain (31a). □

Appendix D: Proof for Lemma 2

Define the convex function

$$f_\alpha(t) = \frac{1}{\alpha-1}(t^\alpha - 1), \tag{113}$$

then for the two distributions P and Q over the support $\mathcal{Y}$, we have a f-divergence $\mathcal{H}_\alpha(P\|Q)$, which is the Hellinger divergence of order α [51], given by

$$\mathcal{H}_\alpha(P\|Q) = \frac{1}{\alpha-1}\left(\sum_{\mathcal{Y}} P(y)^\alpha Q(y)^{1-\alpha} - 1\right). \tag{114}$$

Therefore, the Rényi divergence can be written in terms of the Hellinger divergence as

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1}\log(1+(\alpha-1)\mathcal{H}_\alpha(P\|Q)). \tag{115}$$

Thus, since $z \mapsto \frac{1}{\alpha-1}\log(1+(\alpha-1)z)$ is monotonically increasing in z for $\alpha > 1$, we can write maximal α-leakage as

$$\begin{aligned}
\mathcal{L}_\alpha^{\max}(X\to Y) &= \sup_{P_X}\inf_{Q_Y} D_\alpha(P_{X,Y}\|P_X\times Q_Y) && (116)\\
&= \frac{1}{\alpha-1}\log\left(1+(\alpha-1)\sup_{P_X}\inf_{Q_Y}\mathcal{H}_\alpha(X\to Y)\right) && (117)\\
&= \frac{1}{\alpha-1}\log\left(1+(\alpha-1)\mathcal{L}_{\mathcal{H}_\alpha}(X\to Y)\right). && (118)
\end{aligned}$$

That is, for $\alpha > 1$ maximal α-leakage is a monotonic function of the Hellinger divergence-based measure. □

Appendix E: Proof of Theorem 3

The proof of part 1: We know that for $\alpha \geq 1$, $I_\alpha^{\mathrm{S}}(X;Y)$ is quasi-convex $P_{Y|X}$ for given P_X [56, Thm. 2.7.4], [48, Thm. 10]. In addition, the supremum of a set of quasi-convex functions is also quasi-convex, i.e., if the function $f(a,b)$ is quasi-convex in b for any given a, the supremum $\sup_a f(a,b)$ is also quasi-convex in b [52]. Therefore, maximal α-leakage in (31) is quasi-convex $P_{Y|X}$.

The proof of part 2: Let $\beta > \alpha \geq 1$, and $P_{X\alpha}^\star = \arg\sup_{P_X} I_\alpha^{\mathrm{S}}(P_X, P_{Y|X})$ for given $P_{Y|X}$, such that

$$\begin{aligned}
\mathcal{L}_\alpha^{\max}(X\to Y) &= I_\alpha^{\mathrm{S}}(P_{X\alpha}^\star, P_{Y|X}) && (119)\\
&\leq I_\beta^{\mathrm{S}}(P_{X\alpha}^\star, P_{Y|X}) && (120)\\
&\leq \sup_{P_X} I_\beta^{\mathrm{S}}(P_X, P_{Y|X}) && (121)\\
&= \mathcal{L}_\beta^{\max}(X\to Y) && (122)
\end{aligned}$$

where (120) results from that I_α^{S} is non-decreasing in α for $\alpha > 0$ [48, Thm. 4], and the equality in (121) holds if and only if $P_{X\alpha}^\star = \arg\sup_{P_X} I_\beta(P_X, P_{Y|X})$.

The proof of part 3: Let random variables X, Y and Z form the Markov chain $X - Y - Z$. Making use of that Sibson MI of order $\alpha > 1$ satisfies data processing inequalities [43, Thm. 3], i.e.,

$$\begin{aligned}
I_\alpha^{\mathrm{S}}(X;Z) &\leq I_\alpha^{\mathrm{S}}(X;Y) && (123)\\
I_\alpha^{\mathrm{S}}(X;Z) &\leq I_\alpha^{\mathrm{S}}(Y;Z), && (124)
\end{aligned}$$

we prove that maximal α-leakage satisfies data processing inequalities as follows.

We first prove (41a). Let $P_X^\star = \arg\sup_{P_X} I_\alpha^{\mathrm{S}}(P_X, P_{Z|X})$. For the Markov chain $X - Y - Z$, we have

$$\begin{aligned}
\mathcal{L}_\alpha^{\max}(X\to Z) &= I_\alpha^{\mathrm{S}}(P_X^\star, P_{Z|X}) && (125)\\
&\leq I_\alpha^{\mathrm{S}}(P_X^\star, P_{Y|X}) && (126)\\
&\leq \sup_{P_X} I_\alpha^{\mathrm{S}}(P_X, P_{Y|X}) && (127)\\
&= \mathcal{L}_\alpha^{\max}(X\to Y) && (128)
\end{aligned}$$

where the inequality in (126) results from (123). Similarly, the inequality in (41b) can be proved directly from (124).

The proof of part 4: For $\alpha = 1$, we have

$$\mathcal{L}_1^{\max}(X\to Y) = I(X;Y) \geq 0, \tag{129}$$

with equality if and only if X is independent of Y [56]. For $1<\alpha\leq\infty$, referring to (6) and (31a) we have

$$\begin{aligned}&\mathcal{L}_\alpha^{\max}(X\to Y)\\ =&\sup_{P_X}\frac{\alpha}{\alpha-1}\log\sum_y\left(\sum_x P_X(x)P_{Y|X}(y|x)^\alpha\right)^{\frac{1}{\alpha}} &(130)\\ \geq&\sup_{P_X}\frac{\alpha}{\alpha-1}\log\sum_y\left(\sum_x P_X(x)P_{Y|X}(y|x)\right)^{\frac{\alpha}{\alpha}} &(131)\\ =&\sup_{P_X}\frac{\alpha}{\alpha-1}\log 1=0, &(132)\end{aligned}$$

where (131) results from applying Jensen's inequality to the convex function $f:t\to t^\alpha$ $(t\geq 0)$, such that the equality holds if and only if given any $y\in\mathcal{Y}$, $P_{Y|X}(y|x)$ are the same for all $x\in\mathcal{X}$, such that

$$P_{Y|X}(y|x)=P_Y(y)\quad x\in\mathcal{X},y\in\mathcal{Y} \tag{133}$$

which means X and Y are independent, i.e., $P_{Y|X}$ is a rank-1 row stochastic matrix.

For $\alpha=1$, from (31b) we know $\mathcal{L}_1^{\max}(X\to Y)=I(X;Y)$. Therefore,

$$\begin{aligned}&\mathcal{L}_1^{\max}(X\to Y)-H(X)\\ =&\sum_{x\in\mathcal{X},y\in\mathcal{Y}}P(x,y)\log\frac{P(y|x)}{P(y)}-\sum_{x\in\mathcal{X}}P(x)\log\frac{1}{P(x)} &(134)\\ =&\sum_{x,y}P(x,y)\log\frac{P(y|x)}{P(y)}-\sum_{x,y}P(x,y)\log\frac{1}{P(x)} &(135)\\ =&\sum_{x,y}P(x,y)\log P(x|y)\leq 0, &(136)\end{aligned}$$

with equality if and only if for all $x,y\in\mathcal{X}\times\mathcal{Y}$, the conditional probability $P_{X|Y}(x|y)$ is either 1 or 0. That is, $\mathcal{L}_1^{\max}(X\to Y)\leq H(X)$ with equality if and only if X is a deterministic function of Y [57, Lem. 1]. For $1<\alpha\leq\infty$, from the monotonicity of maximal α-leakage in α and (31a), we have

$$\begin{aligned}\mathcal{L}_\alpha^{\max}(X\to Y)&\leq\mathcal{L}_\infty^{\max}(X\to Y) &(137)\\ &=\log\sum_{y\in\mathcal{Y}}\max_x P_{Y|X}(y|x) &(138)\\ &\leq\log\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}}P_{Y|X}(y|x)=\log|\mathcal{X}|. &(139)\end{aligned}$$

where the equality in (139) holds if and only if for every $y\in\mathcal{Y}$, $\sum_{\mathcal{X}}P(y|x)=\max_x P(y|x)$, i.e., X is a deterministic function of Y. To prove that for $\alpha\in(1,\infty)$, the upper bound in (139) is achievable, we construct a mapping $P_{X\Leftarrow Y}$ such that X is a deterministic function of Y. That is, for every $y\in\mathcal{Y}$, there exists a unique $x_y\in\mathcal{X}$ such that $P(x_y|y)=1$. Therefore, we have $x_y=\arg_x P_{X\Leftarrow Y}(y|x)>0$. For $\alpha\in(1,\infty)$, from (6) and (31b) we have

$$\begin{aligned}&\mathcal{L}_\alpha^{\max}(P_{X\Leftarrow Y})\\ =&\sup_{P_X}\frac{\alpha}{\alpha-1}\log\sum_{y\in\mathcal{Y}}\left(P_X^{\frac{1}{\alpha}}(x_y)P_{X\Leftarrow Y}(y|x_y)\right) &(140)\\ =&\sup_{P_X}\frac{\alpha}{\alpha-1}\log\sum_{x\in\mathcal{X}}P_X^{\frac{1}{\alpha}}(x); &(141)\end{aligned}$$

in addition, since the function maximized in (141) is symmetric and concave in P_X, it is Schur-concave in P_X, and therefore, the optimal distribution of X achieving the supreme in (141) is uniform. Thus,

$$\mathcal{L}_\alpha^{\max}(P_{X\Leftarrow Y})=\log|\mathcal{X}|,\quad 1<\alpha\leq\infty. \tag{142}$$

Therefore, maximal α-leakage achieves its maximal value $\log|\mathcal{X}|$ and $H(P_X)$ for $\alpha>1$ and $\alpha=1$, respectively, if and only if X is a deterministic function of Y. □

Appendix F: Proof for Theorem 4

To prove Thm. 4, we define a divergence function k_α for $\alpha>1$ and provide a lower bound for its sum in the following definition and lemma, respectively.

Definition 8. *Given two discrete distributions P_Y and Q_Y over the support $\mathcal{Y}$, a divergence function k_α for $\alpha>1$ is defined as*

$$k_\alpha(P_Y\|Q_Y)\triangleq\sum_y Q_Y(y)\left(\frac{P_Y(y)}{Q_Y(y)}\right)^\alpha. \tag{143}$$

Proposition 1. *The function $k_\alpha(P_Y\|Q_Y)$ in (143) is jointly convex in (P_Y,Q_Y), and $k_\alpha(P_Y\|Q_Y)\geq 1$ with equality if and only if $P_Y=Q_Y$.*

Proof. For $\alpha\geq 1$, the function $f(t)=t^\alpha$ is convex in $t\geq 0$, such that the perspective of $f(t)$, defined as $g(t,a)=af(t/a)$ $(a>0)$, is convex in (t,a) [52, Chapter. 3.2.6]. Let $t=P_Y(y)$ and $a=Q_Y(y)>0$ such that the perspective function can be written as

$$g(P_Y(y),Q_Y(y))=Q_Y(y)\left(\frac{P_Y(y)}{Q_Y(y)}\right)^\alpha, \tag{144}$$

which is therefore convex in $((P_Y(y),Q_Y(y))$. For $Q_Y(y)=0$, the function $g(P_Y(y),Q_Y(y))$ is zero, which is also convex in $((P_Y(y),Q_Y(y))$. Thus, the function $k_\alpha(P_Y\|Q_Y)$ in (143) is a sum of convex functions, and therefore, it is convex in $(P_Y(y),Q_Y(y))$.

Let $t=\frac{P_Y(y)}{Q_Y(y)}$. From the convexity of $f(t)=t^\alpha$ in $t\geq 0$ and Jensen's inequality [52, Chapter. 3.1.8], we have that

$$\begin{aligned}k_\alpha(P_Y\|Q_Y)&\triangleq\sum_y Q_Y(y)\left(\frac{P_Y(y)}{Q_Y(y)}\right)^\alpha &(145)\\ &\geq\sum_y\left(\sum_y Q_Y(y)\frac{P_Y(y)}{Q_Y(y)}\right)^\alpha=1. &(146)\end{aligned}$$

□

Lemma 3. *Let K be a positive integer with $K<\infty$. Given a group of distributions $\{P_k:k\in[1,K]\}$ and an arbitrary*

distribution P on a discrete set $\mathcal{Y}$, there is

$$\sum_{k=1}^{K} k_\alpha(P_k\|P) \geq \sum_{k=1}^{K} k_\alpha(P_k\|P_c) \tag{147}$$

$$= \left(\sum_y \left(\sum_{k=1}^{K} P_k(y)^\alpha\right)^{\frac{1}{\alpha}}\right)^\alpha, \tag{148}$$

with equality if and only if $P = P_c$, where P_c is given by

$$P_c(y) = \frac{1}{Z}\left(\sum_{k=1}^{K} P_k(y)^\alpha\right)^{\frac{1}{\alpha}}, \ \alpha \in [1,\infty] \tag{149}$$

where Z is the constant as

$$Z = \sum_y \left(\sum_{k=1}^{K} P_k(y)^\alpha\right)^{\frac{1}{\alpha}}, \tag{150}$$

which guarantees that P_c is a distribution.

Proof. From the definition k_α in (143), we have

$$\begin{aligned}
&\sum_{k=1}^{K} k_\alpha(P_k\|P) - \sum_{k=1}^{K} k_\alpha(P_k\|P_c)\\
=&\sum_{k=1}^{K}\sum_y P_k(y)^\alpha \left(P(y)^{1-\alpha} - P_c(y)^{1-\alpha}\right) &(151)\\
=&\sum_y \left(\sum_{k=1}^{K} P_k(y)^\alpha\right)\left(P(y)^{1-\alpha} - P_c(y)^{1-\alpha}\right) &(152)\\
=&\sum_y Z^\alpha P_c(y)^\alpha \left(P(y)^{1-\alpha} - P_c(y)^{1-\alpha}\right) &(153)\\
=&Z^\alpha \sum_y \left(P_c(y)^\alpha P(y)^{1-\alpha} - P_c(y)\right) &(154)\\
=&Z^\alpha(k_\alpha(P_c\|P) - 1) \geq 0 &(155)
\end{aligned}$$

with equality if and only if $P = P_c$. In addition, making use of the expression of P_c and Z in (149) and (150), respectively, we have

$$\begin{aligned}
&\sum_{k=1}^{K} k_\alpha(P_k\|P_c)\\
=&\sum_{k=1}^{K}\sum_y P_c(y)\left(\frac{P_k(y)}{P_c(y)}\right)^\alpha &(156)\\
=&\sum_{k=1}^{K}\sum_y Z^{\alpha-1}\left(\sum_{k'=1}^{K} P_{k'}(y)^\alpha\right)^{\frac{1}{\alpha}} \frac{P_k(y)^\alpha}{\sum\limits_{k'=1}^{K} P_{k'}(y)^\alpha} &(157)\\
=&Z^{\alpha-1}\sum_y \left(\sum_{k'=1}^{K} P_{k'}(y)^\alpha\right)^{\frac{1}{\alpha}} \frac{\sum_{k=1}^{K} P_k(y)^\alpha}{\sum_{k'=1}^{K} P_{k'}(y)^\alpha} &(158)\\
=&\left(\sum_y \left(\sum_{k=1}^{K} P_k(y)^\alpha\right)^{\frac{1}{\alpha}}\right)^\alpha. &(159)
\end{aligned}$$

□

Making use of the results in Lemma 3, we prove Thm. 4 as follows.

Proof. From Thm. 2, we have that for $\alpha > 1$

$$\begin{aligned}
&\mathcal{L}_\alpha^{\max}(X \to Y)\\
=&\sup_{P_{\tilde{X}}} I_\alpha^{\mathrm{S}}(\tilde{X}, Y) = \sup_{P_{\tilde{X}}}\inf_{Q_Y} D_\alpha(P_{\tilde{X}}P_{Y|X}\|P_{\tilde{X}}Q_Y) &(160)\\
=&\sup_{P_{\tilde{X}}}\inf_{Q_Y} \frac{1}{\alpha-1}\log\sum_x P_{\tilde{X}}(x)k_\alpha(P_{Y|X=x}\|Q_Y). &(161)
\end{aligned}$$

For $\alpha > 1$, the function $f : t \to \frac{1}{\alpha-1}\log t$ is increasing in $t \geq 0$. Therefore, we simplify the optimization in (161) as

$$\sup_{P_{\tilde{X}}}\inf_{Q_Y}\sum_x P_{\tilde{X}}(x)k_\alpha(P_{Y|X=x}\|Q_Y) \tag{162}$$

and provide a lower bound of (162) as follows. Since the divergence function k_α is joint convex in the pair of distributions, the objective function in (162) is joint convex in $(P_{Y|X}, Q_Y)$ for fixed $P_{\tilde{X}}$, and linear in $P_{\tilde{X}}$ for fixed $(P_{Y|X}, Q_Y)$. Therefore, the max-min equals to the min-max as followed:

$$\begin{aligned}
&\sup_{P_{\tilde{X}}}\inf_{Q_Y}\sum_x P_{\tilde{X}}(x)k_\alpha(P_{Y|X=x}\|Q_Y)\\
=&\inf_{Q_Y}\sup_{P_{\tilde{X}}}\sum_x P_{\tilde{X}}(x)k_\alpha(P_{Y|X=x}\|Q_Y) &(163)\\
=&\inf_{Q_Y}\max_x k_\alpha(P_{Y|X=x}\|Q_Y) &(164)\\
\geq&\inf_{Q_Y}\frac{\sum_x k_\alpha(P_{Y|X=x}\|Q_Y)}{|\mathcal{X}|} &(165)\\
\geq&\frac{\sum_x k_\alpha(P_{Y|X=x}\|P_c)}{|\mathcal{X}|} &(166)\\
=&\frac{1}{|\mathcal{X}|}\left(\sum_y \|P_{Y|X}(y|\cdot)\|_\alpha\right)^\alpha, &(167)
\end{aligned}$$

where the inequality in (166) is directly from (147) in Lemma 3 with equality if and only if

$$Q_Y(y) = P_c(y) = \frac{1}{Z}\|P_{Y|X}(y|\cdot)\|_\alpha, \tag{168}$$

with the constant $Z = \sum_y \|P_{Y|X}(y|\cdot)\|_\alpha$. Therefore, for any $P_{Y|X}$, we have

$$\mathcal{L}_\alpha^{\max}(X \to Y) \geq \frac{\alpha}{\alpha-1}\log\frac{\sum_y \|P_{Y|X}(y|\cdot)\|_\alpha}{|\mathcal{X}|^{\frac{1}{\alpha}}}, \tag{169}$$

with equality if and only if the $P_{Y|X}$ guarantees that the divergence function $k_\alpha(P_{Y|X=x}\|P_c)$ are the same for all $x \in \mathcal{X}$, i.e., the $P_{Y|X}$ satisfies (46). □

Appendix G: Proof of Theorem 5

Let $\mathcal{Y}_1$ and $\mathcal{Y}_2$ be the alphabets of Y_1 and Y_2, respectively. For any $(y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2$, due to the Markov chain $Y_1 - X - Y_2$, the corresponding entry of the conditional probability matrix of (Y_1, Y_2) given X is

$$P(y_1y_2|x) = P(y_1|x)P(y_2|xy_1) = P(y_1|x)P(y_2|x).$$

Therefore, for $\alpha \in (1,\infty)$

$$\mathcal{L}_\alpha^{\max}(X \to Y_1, Y_2)$$

$$=\sup_{P_X}\frac{\alpha}{\alpha-1}\log\sum_{y_1,y_2}\left(\sum_x P_X(x)P_{Y_1,Y_2|X}(y_1,y_2|x)^\alpha\right)^{\frac{1}{\alpha}} \quad (170)$$

$$=\sup_{P_X}\frac{\alpha}{\alpha-1}\log\sum_{y_1,y_2}\left(\sum_x P(x)P(y_1|x)^\alpha P(y_2|x)^\alpha\right)^{\frac{1}{\alpha}}. \quad (171)$$

Let $K(y_1)=\sum_{x\in\mathcal{X}}P_X(x)P_{Y_1|X}(y_1|x)^\alpha$, for all $y_1\in\mathcal{Y}_1$, such that we can construct a set of distributions over $\mathcal{X}$ as

$$P_{\tilde{X}}(x|y_1)=\frac{P_X(x)P_{Y_1|X}(y_1|x)^\alpha}{K(y_1)}. \quad (172)$$

Therefore, from (171), $\mathcal{L}_\alpha^{\max}(X\to Y_1,Y_2)$ can be rewritten as

$$\begin{aligned}
&\mathcal{L}_\alpha^{\max}(X\to Y_1,Y_2)\\
&=\sup_{P_X}\frac{\alpha}{\alpha-1}\log\sum_{y_1,y_2\in\mathcal{Y}_1\times\mathcal{Y}_2}\left(\sum_{x\in\mathcal{X}}K(y_1)P_{\tilde{X}}(x|y_1)P_{Y_2|X}(y_2|x)^\alpha\right)^{\frac{1}{\alpha}} && (173)\\
&=\sup_{P_X}\frac{\alpha}{\alpha-1}\log\sum_{y_1,y_2}\left(\left(\sum_x P_X(x)P_{Y_1|X}(y_1|x)^\alpha\right)^{\frac{1}{\alpha}}\cdot\left(\sum_x P_{\tilde{X}}(x|y_1)P_{Y_2|X}(y_2|x)^\alpha\right)^{\frac{1}{\alpha}}\right) && (174)\\
&=\sup_{P_X}\frac{\alpha}{\alpha-1}\log\sum_{y_1}\left(\left(\sum_x P_X(x)P_{Y_1|X}(y_1|x)^\alpha\right)^{\frac{1}{\alpha}}\cdot\sum_{y_2}\left(\sum_x P_{\tilde{X}}(x|y_1)P_{Y_2|X}(y_2|x)^\alpha\right)^{\frac{1}{\alpha}}\right) && (175)\\
&\le\sup_{P_X}\frac{\alpha}{\alpha-1}\log\left(\sum_{y_1}\left(\sum_x P_X(x)P_{Y_1|X}(y_1|x)^\alpha\right)^{\frac{1}{\alpha}}\cdot\max_{y_1}\sum_{y_2}\left(\sum_x P_{\tilde{X}}(x|y_1)P_{Y_2|X}(y_2|x)^\alpha\right)^{\frac{1}{\alpha}}\right) && (176)\\
&=\sup_{P_X}\frac{\alpha}{\alpha-1}\log\left(\sum_{y_1}\left(\sum_x P_X(x)P_{Y_1|X}(y_1|x)^\alpha\right)^{\frac{1}{\alpha}}\cdot\sum_{y_2}\left(\sum_x P_{\tilde{X}}(x|y_1^\star)P_{Y_2|X}(y_2|x)^\alpha\right)^{\frac{1}{\alpha}}\right) && (177)(178)\\
&\le\sup_{P_X}\frac{\alpha}{\alpha-1}\log\sum_{y_1}\left(\sum_x P_X(x)P_{Y_1|X}(y_1|x)^\alpha\right)^{\frac{1}{\alpha}}\\
&\quad+\sup_{P_{\tilde{X}}}\frac{\alpha}{\alpha-1}\log\sum_{y_2}\left(\sum_x P_{\tilde{X}}(x)P_{Y_2|X}(y_2|x)^\alpha\right)^{\frac{1}{\alpha}} && (179)\\
&=\mathcal{L}_\alpha^{\max}(X\to Y_1)+\mathcal{L}_\alpha^{\max}(X\to Y_2), && (180)
\end{aligned}$$

where $y_1^\star$ in (178) is the optimal y_1 achieving the maximum in (176). Therefore, the equality in (176) holds if and only if for all $y_1\in\mathcal{Y}_1$

$$\sum_{y_2}\left(\sum_x P_{\tilde{X}}(x|y_1)P_{Y_2|X}(y_2|x)^\alpha\right)^{\frac{1}{\alpha}}$$

$$=\sum_{y_2}\left(\sum_x P_{\tilde{X}}(x|y_1^\star)P_{Y_2|X}(y_2|x)^\alpha\right)^{\frac{1}{\alpha}}; \quad (181)$$

and the equality in (179) holds if and only if the optimal solutions $P_X^\star$ and $P_{\tilde{X}}^\star$ of the two maximizations in (179) satisfy, for all $x\in\mathcal{X}$,

$$P_{\tilde{X}}^\star(x)=\frac{P_X^\star(x)P_{Y_1|X}^\alpha(y_1^\star|x)}{\sum_{x\in\mathcal{X}}P_X(x)P_{Y_1|X}^\alpha(y_1^\star|x)}. \quad (182)$$

Now we consider $\alpha=1$. For Y_1-X-Y_2, we have

$$I(Y_2;X|Y_1)\le I(Y_2;X). \quad (183)$$

From Thm. 2, there is

$$\begin{aligned}
&\mathcal{L}_1^{\max}(X\to Y_1,Y_2)\\
&=I(X;Y_1)+I(X;Y_2|Y_1) && (184)\\
&\le I(X;Y_1)+I(X;Y_2) && (185)\\
&=\mathcal{L}_1^{\max}(X\to Y_1)+\mathcal{L}_1^{\max}(X\to Y_2). && (186)
\end{aligned}$$

For $\alpha=\infty$, we also have

$$\begin{aligned}
&\mathcal{L}_\infty^{\max}(X\to Y_1,Y_2)\\
&=\log\sum_{y_1,y_2\in\mathcal{Y}_1\times\mathcal{Y}_2}\max_{x\in\mathcal{X}}P(y_1|x)P(y_2|x) && (187)\\
&\le\log\sum_{y_1,y_2\in\mathcal{Y}_1\times\mathcal{Y}_2}\left(\max_{x\in\mathcal{X}}P(y_1|x)\right)\left(\max_{x\in\mathcal{X}}P(y_2|x)\right) && (188)\\
&=\log\sum_{y_1\in\mathcal{Y}_1}\max_{x\in\mathcal{X}}P(y_1|x)+\log\sum_{y_2\in\mathcal{Y}_2}\max_{x\in\mathcal{X}}P(y_2|x) && (189)\\
&=\mathcal{L}_\infty^{\max}(X\to Y_1)+\mathcal{L}_\infty^{\max}(X\to Y_2). && (190)
\end{aligned}$$

□

Appendix H: Proof of Theorem 6

For $\alpha>1$, a function $f(t)=\frac{\alpha}{\alpha-1}\log t$ is monotonically increasing in $t>0$. Therefore, to solve maximal α-leakage from X^n to Y^n, i.e.,

$$\begin{aligned}
&\mathcal{L}_\alpha^{\max}(X^n\to Y^n)\\
&=\sup_{P_{\tilde{X}^n}}\frac{\alpha}{\alpha-1}\log\sum_{y^n}\left(\sum_{x^n}P(x^n)P(y^n|x^n)^\alpha\right)^{\frac{1}{\alpha}}, && (191)
\end{aligned}$$

it is sufficient to prove that

$$\begin{aligned}
&\sup_{P_{\tilde{X}^n}}\sum_{y^n}\left(\sum_{x^n}P(x^n)P(y^n|x^n)^\alpha\right)^{\frac{1}{\alpha}}\\
&=\sup_{\substack{P_{\tilde{X}_i}\\ i\in[1,n]}}\prod_{i=1}^n\left(\sum_{y_i}\left(\sum_{x_i}P\left(x_i\right)P\left(y_i|x_i\right)^\alpha\right)^{\frac{1}{\alpha}}\right). && (192)
\end{aligned}$$

For a memoryless $P_{Y^n|X^n}$ with no feedback, we simplify (192) as

$$\sup_{P_{\tilde{X}^n}}\sum_{y^n}\left(\sum_{x^n}\frac{P(x^n,y^n)}{P(y^n|x^n)^{1-\alpha}}\right)^{\frac{1}{\alpha}}$$

$$= \sup_{\prod_{i=1}^n P_{\tilde{X}_i|\tilde{X}_{i-1},\cdots,\tilde{X}_1}} \sum_{y_1,\cdots,y_n} \left(\sum_{x_1,\cdots,x_n} \prod_{i=1}^n \frac{P(y_i,x_i|x_{i-1},y_{i-1},\cdots,x_1y_1)}{P\left(y_i|x^n y_{i-1},\cdots,y_1\right)^{1-\alpha}} \right)^{\frac{1}{\alpha}} \tag{193}$$

$$= \sup_{\prod_{i=1}^n P_{\tilde{X}_i|\tilde{X}_{i-1},\cdots,\tilde{X}_1}} \sum_{y_1,\cdots,y_n} \left(\sum_{x_1,\cdots,x_n} \prod_{i=1}^n \left(\frac{P\left(y_i|x_i,\cdots,x_1\right) P\left(x_i|x_{i-1},\cdots,x_1\right)}{P\left(y_i|x^n\right)^{1-\alpha}} \right) \right)^{\frac{1}{\alpha}} \tag{194}$$

$$= \sup_{\prod_{i=1}^n P_{\tilde{X}_i|\tilde{X}_{i-1},\cdots,\tilde{X}_1}} \sum_{y_1,\cdots,y_n} \left(\sum_{x_1,\cdots,x_n} \prod_{i=1}^n P\left(y_i|x_i\right)^\alpha P\left(x_i|x_{i-1},\cdots,x_1\right) \right)^{\frac{1}{\alpha}} \tag{195}$$

$$= \sup_{\prod_{i=1}^n P_{\tilde{X}_i}} \sum_{y_1,\cdots,y_n} \left(\sum_{x_1,\cdots,x_n} \prod_{i=1}^n P\left(y_i|x_i\right)^\alpha P\left(x_i\right) \right)^{\frac{1}{\alpha}} \tag{196}$$

$$= \sup_{P_{\tilde{X}_1},\cdots,P_{\tilde{X}_n}} \sum_{y_1,\cdots,y_n} \left(\prod_{i=1}^n \sum_{x_i} P\left(x_i\right) P\left(y_i|x_i\right)^\alpha \right)^{\frac{1}{\alpha}} \tag{197}$$

$$= \sup_{\substack{P_{\tilde{X}_i} \\ i\in[1,n]}} \prod_{i=1}^n \left(\sum_{y_i} \left(\sum_{x_i} P\left(x_i\right) P\left(y_i|x_i\right)^\alpha \right)^{\frac{1}{\alpha}} \right) \tag{198}$$

$$= \sup_{P_{\tilde{X}_i}, i\in[1,n]} \prod_{i=1}^n \exp\left\{ \frac{\alpha-1}{\alpha} I_\alpha^{\mathrm{S}}(\tilde{X}_i;Y_i) \right\} \tag{199}$$

where

- (193) is from the chain rule of probability;
- (194) and (195) are directly from the mechanism has no feedback and is memoryless, respectively;
- the equality in (196) holds for memoryless sources, i.e., $P_{\tilde{X}_i|\tilde{X}_{i-1},\cdots,\tilde{X}_1} = P_{\tilde{X}_i}$ for all $i \in [1,n]$;
- both (197) and (198) are from the distributive property of multiplication;
- (199) is from the definition of Sibson MI in (6) and that the base of the logarithm is 2.

Therefore, we have for $\alpha > 1$,

$$\sup_{P_{\tilde{X}^n}} I_\alpha^{\mathrm{S}}(\tilde{X}^n;Y^n) = \sum_{i=1}^n \sup_{P_{\tilde{X}_i}} I_\alpha^{\mathrm{S}}(\tilde{X}_i;Y_i). \tag{200}$$

That is,

$$\mathcal{L}_\alpha^{\max}\left(X^n \to Y^n\right) = \sum_{i=1}^n \mathcal{L}_\alpha^{\max}\left(X_i \to Y_i\right). \tag{201}$$

For $\alpha = 1$, we have

$$I\left(X^n;Y^n\right) = \sum_{i,j=1}^n I\left(X_i;Y_j\middle|X_{i-1},\cdots,X_1,Y_{j-1},\cdots,Y_1\right) \tag{202}$$

$$= \sum_{i,j=1}^n I\left(X_i;Y_j\middle|X_{i-1},\cdots,X_1\right) \tag{203}$$

$$= \sum_{i=1}^n I\left(X_i;Y_i\middle|X_{i-1},\cdots,X_1\right) \tag{204}$$

$$\le \sum_{i=1}^n I\left(X_i;Y_i\right) \tag{205}$$

where

- (202) is from the chain rule of MI;
- (203) and (204) are from the facts that the mechanism has no feedback and is memoryless, respectively;
- from [56, (2.122)], we know that for a Markov chain $X - Y - Z$, conditioning reduces mutual information, i.e., $I(X;Y|Z) \le I(X;Y)$ with equality if and only if $I(X;Z) = 0$. Therefore, since for any $i \in [1,n]$ $(X_{i-1},\cdots,X_1) - X_i - Y_i$, the equality in (196) holds if and only if the source is memoryless, i.e., $P_{\tilde{X}_i|\tilde{X}_{i-1},\cdots,\tilde{X}_1} = P_{\tilde{X}_i}$ for all $i \in [1,n]$.

□

Appendix I: Proof of Theorem 7

Given P_X, the collection of stochastic matrices is denoted as $\mathcal{P}_{Y|X}$. The feasible ball $B_D(x)$ around x is defined in (51). For the distribution dependent PUT in (53), we have

$$\mathrm{PUT}_{\mathrm{HD},\mathcal{L}_f}(D)$$

$$= \inf_{\substack{P_{Y|X}\in\mathcal{P}_{Y|X} \\ :d(X,Y)\le D}} \inf_{Q_Y} D_f(P_{Y|X}P_X\|P_X\times Q_Y) \tag{206}$$

$$= \inf_{Q_Y} \inf_{\substack{P_{Y|X}\in\mathcal{P}_{Y|X} \\ :d(X,Y)\le D}} \sum_{x\in\mathcal{X}} P_X(x) D_f(P_{Y|X=x}\|Q_Y) \tag{207}$$

$$= \inf_{Q_Y} \sum_{x\in\mathcal{X}} P_X(x) \inf_{\substack{P_{Y|X=x} \\ Y\in B_D(x)}} \sum_{y\in\mathcal{Y}} Q_Y(y) f\left(\frac{P_{Y|X}(y|x)}{Q_Y(y)}\right) \tag{208}$$

$$= \inf_{Q_Y} \sum_{x\in\mathcal{X}} P_X(x) \inf_{\substack{P_{Y|X=x} \\ Y\in B_D(x)}} \left(\sum_{y\in B_D(x)^c} Q_Y(y) f\left(\frac{P_{Y|X}(y|x)}{Q_Y(y)}\right) + \frac{Q_Y(B_D(x))}{Q_Y(B_D(x))} \sum_{y\in B_D(x)} Q_Y(y) f\left(\frac{P_{Y|X}(y|x)}{Q_Y(y)}\right) \right) \tag{209}$$

$$= \inf_{Q_Y} \sum_{x\in\mathcal{X}} P_X(x) \inf_{\substack{P_{Y|X=x} \\ Y\in B_D(x)}} \left(\sum_{y\in B_D(x)^c} Q_Y(y) f\left(0\right) + Q_Y(B_D(x)) \sum_{y\in B_D(x)} \frac{Q_Y(y)}{Q_Y(B_D(x))} f\left(\frac{P_{Y|X}(y|x)}{Q_Y(y)}\right) \right) \tag{210}$$

$$\ge \inf_{Q_Y} \sum_{x\in\mathcal{X}} P_X(x) \inf_{\substack{P_{Y|X=x} \\ Y\in B_D(x)}} \left(Q_Y\left(B_D(x)^c\right) f(0) + Q_Y(B_D(x)) f\left(\frac{1}{Q_Y(B_D(x))}\right) \right) \tag{211}$$

$$= f(0) + \inf_{Q_Y} \sum_{x\in\mathcal{X}} P(x) Q_Y(B_D(x)) \left(f\left(\frac{1}{Q_Y(B_D(x))}\right) - f(0) \right) \tag{212}$$

where

- (207) follows from the fact that $D_f(P_{Y|X}P_X\|P_X\times Q_Y)$ is convex in $(P_{Y|X},Q_Y)$ for fixed P_X,
- (210) is directly from the hard distortion constraint $d(X;Y) \le 0$ such that for any $y \notin B_D(x)$ $P_{Y|X}(y|x) = 0$, and therefore, $\sum_{y\in B_D(x)} P_{Y|X}(y|x) = 1$,

- (211) is from the Jensen's inequality such that

$$\sum_{y\in B_D(x)} \frac{Q_Y(y)}{Q_Y(B_D(x))} f\left(\frac{P_{Y|X}(y|x)}{Q_Y(y)}\right) \geq f\left(\sum_{y\in B_D(x)} \frac{Q_Y(y)}{Q_Y(B_D(x))} \frac{P_{Y|X}(y|x)}{Q_Y(y)}\right) \quad (213)$$

$$= f\left(\frac{\sum\limits_{y\in B_D(x)} P_{Y|X}(y|x)}{Q_Y(B_D(x))}\right) = f\left(\frac{1}{Q_Y(B_D(x))}\right), \quad (214)$$

with equality if and only if there is a mechanism $P_{Y|X}$ satisfying

$$\frac{P_{Y|X}(y|x)}{Q_Y(y)} = \frac{\mathbf{1}(y \in B_D(x))}{Q_Y(B_D(x))}. \quad (215)$$

Note that $f : \mathbb{R}_+ \to \mathbb{R}$ is a convex function, such that the function $tf(\frac{1}{t})$ is convex in $t \in \mathbb{R}_+$. Therefore, the objective function in (212) is convex in Q_Y. Furthermore, in (212) the feasible region of Q_Y is the probability distribution simplex over the set $\{B_D(x), x \in \mathcal{X}\}$. For finite supports $\mathcal{X}$ and $\mathcal{Y}$ of X and Y, respectively, the set $\{B_D(x), x \in \mathcal{X}\}$ is a compact, and therefore, the infimum in (212) is achievable. □

Appendix J: Proof of Theorem 8

Given P_X, the collection of stochastic matrices is denoted as $\mathcal{P}_{Y|X}$. The feasible ball $B_D(x)$ around x is defined in (51). For the distribution independent PUT in (56), we have

$$\mathrm{PUT}_{\mathrm{HD},\mathcal{L}_f^{\max}}(D)$$

$$= \inf_{\substack{P_{Y|X}\in\mathcal{P}_{Y|X}\\ :d(X,Y)\leq D}} \sup_{P_{\tilde{X}}} \inf_{Q_Y} D_f(P_{\tilde{X}}P_{Y|X}\|P_{\tilde{X}}\times Q_Y) \quad (216)$$

$$= \inf_{Q_Y} \sup_{P_{\tilde{X}}} \inf_{\substack{P_{Y|X}\in\mathcal{P}_{Y|X}\\ :d(X,Y)\leq D}} D_f(P_{\tilde{X}}P_{Y|X}\|P_{\tilde{X}}\times Q_Y) \quad (217)$$

$$= \inf_{Q_Y} \sup_{P_{\tilde{X}}} \inf_{\substack{P_{Y|X}\in\mathcal{P}_{Y|X}\\ :d(X,Y)\leq D}} \sum_{x\in\mathcal{X}} P_{\tilde{X}}(x) D_f(P_{Y|X=x}\|Q_Y) \quad (218)$$

$$= \inf_{Q_Y} \sup_{P_{\tilde{X}}} \sum_{x\in\mathcal{X}} P_{\tilde{X}}(x) \inf_{\substack{P_{Y|X=x}\\ Y\in B_D(x)}} \sum_{y\in\mathcal{Y}} Q_Y(y) f\left(\frac{P_{Y|X}(y|x)}{Q_Y(y)}\right) \quad (219)$$

$$= \inf_{Q_Y} \sup_{P_{\tilde{X}}} \sum_{x\in\mathcal{X}} P_{\tilde{X}}(x) \inf_{\substack{P_{Y|X=x}\\ Y\in B_D(x)}} \left(\sum_{y\in B_D(x)} Q_Y(y) \cdot f\left(\frac{P_{Y|X}(y|x)}{Q_Y(y)}\right) + \sum_{y\in B_D(x)^c} Q_Y(y) f(0)\right) \quad (220)$$

$$= \inf_{Q_Y} \sup_{P_{\tilde{X}}} \sum_{x\in\mathcal{X}} P_{\tilde{X}}(x) \inf_{\substack{P_{Y|X=x}\\ Y\in B_D(x)}} \left(Q_Y(B_D(x)) \sum_{y\in B_D(x)} \frac{Q_Y(y)}{Q_Y(B_D(x))} f\left(\frac{P_{Y|X}(y|x)}{Q_Y(y)}\right) + Q_Y(B_D(x)^c) f(0)\right) \quad (221)$$

$$\geq \inf_{Q_Y} \sup_{P_{\tilde{X}}} \sum_{x\in\mathcal{X}} P_{\tilde{X}}(x) \inf_{\substack{P_{Y|X=x}\\ Y\in B_D(x)}} \left(Q_Y(B_D(x)) \cdot f\left(\frac{1}{Q_Y(B_D(x))}\right) + Q_Y(B_D(x)^c) f(0)\right) \quad (222)$$

$$= \inf_{Q_Y} \sup_{P_{\tilde{X}}} \sum_{x\in\mathcal{X}} P_{\tilde{X}}(x) \left(Q_Y(B_D(x)) f\left(\frac{1}{Q_Y(B_D(x))}\right) + \big(1 - Q_Y(B_D(x))\big) f(0)\right) \quad (223)$$

$$= \inf_{Q_Y} \sup_{P_{\tilde{X}}} \sum_{x\in\mathcal{X}} P_{\tilde{X}}(x)\, g\big(Q_Y(B_D(x))\big) \quad (224)$$

$$= \inf_{Q_Y} \sup_{x} g\big(Q_Y(B_D(x))\big) \quad (225)$$

where

- (217) and (219) follow from the fact that $D_f(P_{\tilde{X}}P_{Y|X}\|P_{\tilde{X}} \times Q_Y)$ is linear in $P_{\tilde{X}}$ for fixed $(P_{Y|X}, Q_Y)$ and convex in $(P_{Y|X}, Q_Y)$ for fixed $P_{\tilde{X}}$,
- (222) follows from the convexity of f and Jensen's inequality. The equality holds if and only if there exists a mechanism $P_{Y|X}$ satisfying (215).
- (224) results from $q \triangleq Q_Y(B_D(x))$ and

$$g(q) \triangleq qf(q^{-1}) + (1-q)f(0). \quad (226)$$

Due to the convexity of f, we have $f(q^{-1}) - f(0) \leq f'(q^{-1})\left(q^{-1} - 0\right)$, from which, the derivative $g'(q) = f(q^{-1}) - q^{-1}f'(q^{-1}) - f(0) \leq 0$. Therefore, the function g in (226) is non-increasing, such that (225) is simplified as $g(q^\star)$, where $q^\star$ is given by

$$q^\star \triangleq \sup_{Q_Y} \inf_{x} Q_Y(B_D(x)). \quad (227)$$

Note that in (227), the feasible region of Q_Y is the probability distribution simplex over the set $\{B_D(x), x \in \mathcal{X}\}$. For finite supports $\mathcal{X}$ and $\mathcal{Y}$ of X and Y, respectively, the set $\{B_D(x), x \in \mathcal{X}\}$ is a compact, and therefore, the supremum in (227) is achievable.

□

Appendix K: Proof of Theorem 9

From Thm. 1, we know that for $\alpha \geq 1$, α-leakage $\mathcal{L}_\alpha(S;Y)$ equals to Arimoto MI $I_\alpha^{\mathrm{A}}(S;Y)$. Since $I_\alpha^{\mathrm{A}}(S;Y) = H_\alpha(S) - H_\alpha^{\mathrm{A}}(S|Y)$ and $H_\alpha(S)$ is independent of $P_{Y|S,X}$, to minimize $I_\alpha^{\mathrm{A}}(S;Y)$ with respect to $P_{Y|S,X}$ can be simplified to maximize $H_\alpha^{\mathrm{A}}(S|Y)$. In addition, for $\alpha > 1$, the function $g : t \to \frac{\alpha}{1-\alpha}\log t$ is a monotonically non-increase function in $t > 0$. Therefore, the problem in (61) can be simplified to

$$\inf_{\substack{P_{Y|SX}\\ :d(X,Y)\leq D}} \sum_{y\in\mathcal{Y}} \Big(\sum_{s\in\mathcal{S}} P(s,y)^\alpha\Big)^{\frac{1}{\alpha}}. \quad (228)$$

The hard distortion on X and Y in (61) determines a collection of feasible x and therefore s for each y. We define the two collections for each $y \in \mathcal{Y}$ as

$$\mathcal{X}_D(y) \triangleq \{x \in \mathcal{X} : d(x,y) \leq D\}, \quad (229)$$

$$\mathcal{S}_D(y) \triangleq \{s \in \mathcal{S} : \exists\, x \in \mathcal{X}_D(y), P_{SX}(sx) > 0\}. \quad (230)$$

Note that both sets defined above are independent of the privacy mechanism $P_{Y|S,X}$.

For $\alpha \in (1,\infty)$, we have

$$\inf_{\substack{P_{Y|SX}\\ :d(X,Y)\leq D}} \sum_{y} \Big(\sum_{s} P(s,y)^\alpha\Big)^{\frac{1}{\alpha}}$$

$$= \inf_{\substack{P_{Y|SX}\\ :d(X,Y)\le D}} \sum_{y\in\mathcal{Y}} \left(\sum_{s\in\mathcal{S}} \Big(\sum_{x\in\mathcal{X}} P(y|s,x)P(s,x) \Big)^{\alpha} \right)^{\frac{1}{\alpha}} \tag{231}$$

$$= \inf_{P_{Y|S,X}} \sum_{y} \left(\sum_{\mathcal{S}_D(y)} \left(\sum_{\mathcal{X}_D(y)} P(s,x,y) \right)^{\alpha} + \sum_{\substack{x\notin\mathcal{X}_D(y)\\ s\notin\mathcal{S}_D(y)}} 0 \right)^{\frac{1}{\alpha}} \tag{232}$$

$$= \inf_{P_{Y|S,X}} \sum_{y} \Big(\sum_{s'\in\mathcal{S}_D(y)} P(s')^{\alpha} \Big)^{\frac{1}{\alpha}} \left(\sum_{s\in\mathcal{S}_D(y)} \frac{P(s)^{\alpha}}{\sum_{s'\in\mathcal{S}_D(y)} P(s')^{\alpha}} \Big(\sum_{\mathcal{X}_D(y)} P(x,y|s) \Big)^{\alpha} \right)^{\frac{1}{\alpha}} \tag{233}$$

$$\ge \inf_{P_{Y|S,X}} \sum_{y} \Big(\sum_{s'\in\mathcal{S}_D(y)} P(s')^{\alpha} \Big)^{\frac{1}{\alpha}} \left(\sum_{s\in\mathcal{S}_D(y)} \frac{P(s)^{\alpha}}{\sum_{s'\in\mathcal{S}_D(y)} P(s')^{\alpha}} \Big(\sum_{\mathcal{X}_D(y)} P(x,y|s) \Big) \right) \tag{234}$$

$$= \inf_{P_{Y|S,X}} \sum_{\substack{y,\mathcal{S}_D(y)\\ \mathcal{X}_D(y)}} \Big(\sum_{s'\in\mathcal{S}_D(y)} P(s')^{\alpha} \Big)^{\frac{1}{\alpha}-1} P(s)^{\alpha} P(x,y|s) \tag{235}$$

$$= \inf_{P_{Y|S,X}} \sum_{\substack{s,x\\ B_D(x)}} \Big(\sum_{s'\in\mathcal{S}_D(y)} P(s')^{\alpha} \Big)^{\frac{1}{\alpha}-1} P(s)^{\alpha} P(x,y|s) \tag{236}$$

$$\ge \inf_{P_{Y|S,X}} \sum_{s,x} P(s)^{\alpha} P(x|s) \min_{y\in B_D(x)} \Big(\sum_{s'\in\mathcal{S}_D(y)} P(s')^{\alpha} \Big)^{\frac{1-\alpha}{\alpha}} \tag{237}$$

$$= \sum_{s,x} P(s)^{\alpha} P(x|s) \Big(\max_{y\in B_D(x)} \sum_{s'\in\mathcal{S}_D(y)} P(s')^{\alpha} \Big)^{\frac{1}{\alpha}-1}, \tag{238}$$

where

- (234) is directly from the concavity of the function $g_1 : t \to t^{\frac{1}{\alpha}}$ ($\alpha > 1$) and Jensen's inequality. The equality holds if and only if the optimal $P^{\star}_{Y|S,X}$ achieving the infimum satisfies that for all $s \in \mathcal{S}_D(y)$,
$$P^{\star}(y|s) = \sum_{x\in\mathcal{X}_D(y)} P^{\star}(y|sx)P(x|s) = \frac{P^{\star}_Y(y)}{\sum_{s'\in\mathcal{S}_D(y)} P_S(s')} \tag{239}$$
where $P^{\star}_Y$ is the probability distribution of Y derived from $P^{\star}_{Y|S,X}$.
- in (236), $B_D(x)$ is the feasible ball defined in (51).
- the equality in (237) holds if and only if for any (s,x), all y with $P^{\star}(y|s,x) > 0$ lead to the same $\sum_{s'\in\mathcal{S}_D(y)} P(s')$.
- the equality in (238) is from the fact that the function $g : t \to t^{\frac{1}{\alpha}-1}$ is monotonically non-increasing in $t > 0$ for $\alpha \ge 1$.

Similarly, for $\alpha = \infty$, we have

$$\inf_{\substack{P_{Y|SX}\\ :d(X,Y)\le D}} \sum_{y} P_Y(y) \max_{s} P_{S|Y}(s|y)$$

$$= \inf_{P_{Y|S,X}} \sum_{y} P_Y(y) \max_{\mathcal{S}_D(y)} \Big(\sum_{\mathcal{X}_D(y)} P_{S,X|Y}(s,x|y) \Big) \tag{240}$$

$$\ge \inf_{P_{Y|S,X}} \sum_{y} P(y) \Big(\sum_{\mathcal{S}_D(y)} \frac{P(s)}{\sum_{s'\in\mathcal{S}_D(y)} P(s')} \sum_{\mathcal{X}_D(y)} P(s,x|y) \Big) \tag{241}$$

$$= \inf_{P_{Y|S,X}} \sum_{s,x} \sum_{B_D(x)} \frac{P(s)}{\sum_{s'\in\mathcal{S}_D(y)} P(s')} P(s,x,y) \tag{242}$$

$$\ge \inf_{P_{Y|S,X}} \sum_{s,x} \sum_{B_D(x)} P(s,x,y) \min_{y\in B_D(x)} \frac{P(s)}{\sum_{s'\in\mathcal{S}_D(y)} P(s')} \tag{243}$$

$$= \sum_{s,x} P(s,x)P(s) \left(\max_{y\in B_D(x)} \sum_{s'\in\mathcal{S}_D(y)} P(s') \right)^{-1}. \tag{244}$$

Note that the sufficient and necessary conditions for the equalities in (241) and (243) hold are the same as that for (234) and (237), respectively.

For $\alpha = 1$, $\mathcal{L}_{\alpha}(S \to Y) = I^{\mathrm{A}}(S;Y) = I(S;Y)$, such that

$$\mathrm{PUT}_{\mathrm{HD},\mathcal{L}_{\alpha}}(D) = \inf_{\substack{P_{Y|SX}\\ :d(X,Y)\le D}} \sum_{s,y} P(s,y) \log \frac{P(s,y)}{P(s)P(y)} \tag{245}$$

$$= \inf_{P_{Y|S,X}} \sum_{y} \sum_{\mathcal{S}_D(y)} \left(\Big(\sum_{\mathcal{X}_D(y)} P(s,x,y) \Big) \cdot \log \frac{\sum_{\mathcal{X}_D(y)} P(s,x,y)}{P(s)P(y)} \right) \tag{246}$$

$$\ge \inf_{P_{Y|S,X}} \sum_{y} \left(\Big(\sum_{\mathcal{S}_D(y)} \sum_{\mathcal{X}_D(y)} P(s,x,y) \Big) \cdot \log \frac{\sum_{\mathcal{S}_D(y)} \sum_{\mathcal{X}_D(y)} P(s,x,y)}{\sum_{\mathcal{S}_D(y)} P(s)P(y)} \right) \tag{247}$$

$$= \inf_{P_{Y|S,X}} \sum_{\substack{y,\mathcal{S}_D(y)\\ \mathcal{X}_D(y)}} P(s,x,y) \log \frac{1}{\sum_{s'\in\mathcal{S}_D(y)} P(s')} \tag{248}$$

$$\ge \sum_{s,x} P(s,x) \min_{y\in B_D(x)} \log \frac{1}{\sum_{s'\in\mathcal{S}_D(y)} P(s')}. \tag{249}$$

Note that the inequality in (248) is from log-sum inequality in [56, Thm. 2.7.1], and the sufficient and necessary conditions for the equalities in (248) and (249) hold are the same as that for (234) and (237), respectively. □

Appendix L: Proof of Theorem 10

Define the distortion ball for the type-distance distortion in (65) as

$$B_m(x^n) \triangleq \left\{ y^n : |P_{x^n}(0) - P_{y^n}(0)| \le \frac{m}{n} \right\}. \tag{250}$$

From Corollary 1, to find an optimal mechanism $P^{\star}_{Y^n|X^n}$, we need to find an output distribution $Q^{\star}_{Y^n}$ which optimizes (58) with x^n and y^n in place of x, y.

Note that for the hard distortion $|P_{x^n}(0) - P_{y^n}(0)| \le \frac{m}{n}$, all datasets in a type class share the same group of feasible output datasets, and this feasible group can be represented by output type classes. Therefore, for any $x^n \in T(i)$ ($i \in [0,n]$), we rewrite $B_m(x^n)$ as

$$B_m(x^n) = B_m(T(i)) \triangleq \bigcup_{\substack{|i-j|\le m\\ j\in[0,n]}} T(j). \tag{251}$$

We define a distribution Q_T of type classes for outputs as

$$Q_T(T(j)) \triangleq \sum_{y^n \in T(j)} Q_{Y^n}(y^n), \text{ for } j \in [0,n], \tag{252}$$

such that

$$q^\star = \sup_{Q_T} \min_{i\in[0,n]} Q_T(B_m(T(i))). \tag{253}$$

Note that we replace the infimum by minimum in (253) due to the fact that the infimum over finite integers equals to the minimum. The optimal distribution Q_T is determined by bounding $q*$ from above and below in (253). The upper bound is determined by restricting the optimization in (253) to a judicious choice of a small set of input types. The lower bound is a constructive scheme.

We define an index set $\mathcal{I}_T \subset [0,n]$ for types as

$$I_T \triangleq \left\{ l + (2m+1)k : k \in \left[0, \left\lceil \frac{n+1}{2m+1} \right\rceil - 1\right]\right\} \tag{254}$$

where $l = m$ if $\lceil \frac{n+1}{2m+1} \rceil \le \frac{m+n+1}{2m+1}$, and otherwise, $l = n - \left(\lceil \frac{n+1}{2m+1} \rceil - 1\right)(2m+1)$. From the expression of $\mathcal{I}_T$ in (254), we observe that: (i) the difference between adjacent elements is $2m+1$; (ii) for the first and last elements,

- if $\lceil \frac{n+1}{2m+1} \rceil \le \frac{m+n+1}{2m+1}$ holds, the first element is m and the last element is

$$\begin{aligned} &m + (2m+1)\left(\left\lceil \frac{n+1}{2m+1}\right\rceil - 1\right) \\ =&(2m+1)\left\lceil \frac{n+1}{2m+1}\right\rceil - m - 1 \in [n-m, n], \end{aligned} \tag{255}$$

  due to the inequalities $\frac{n+1}{2m+1} \le \lceil \frac{n+1}{2m+1} \rceil \le \frac{m+n+1}{2m+1}$;
- if $\lceil \frac{n+1}{2m+1} \rceil > \frac{m+n+1}{2m+1}$ holds, the last element is n and the first element is

$$\begin{aligned} &n - \left(\left\lceil \frac{n+1}{2m+1}\right\rceil - 1\right)(2m+1) \\ =&n + 2m + 1 - \left\lceil \frac{n+1}{2m+1}\right\rceil (2m+1) \in [0, m), \end{aligned} \tag{256}$$

  due to the inequalities $\frac{n+1}{2m+1} + 1 - \frac{1}{2m+1} \ge \lceil \frac{n+1}{2m+1} \rceil > \frac{m+n+1}{2m+1}$ for $n \in \mathbb{Z}_{++}$.

Therefore, it is not difficult to see that feasible balls of input type classes indexed by I_T are a partition of the set of all type classes, i.e.,

$$B_m(T(i_1)) \cap B_m(T(i_2)) = \emptyset \quad i_1, i_2 \in \mathcal{I}_T, \tag{257a}$$

$$\bigcup_{j\in[0,n]} T(j) = \bigcup_{i\in\mathcal{I}_T} B_m(T(i)). \tag{257b}$$

Therefore, the problem in (253) is bounded from above by

$$\begin{aligned} q^\star \le& \sup_{Q_T} \min_{i\in\mathcal{I}_T} Q_T(B_m(T(i))) && (258)\\ \le& \sup_{Q_T} \frac{1}{|\mathcal{I}_T|} \sum_{i\in\mathcal{I}_T} Q_T(B_m(T(i))) && (259)\\ =& \sup_{Q_T} \left(\left\lceil \frac{n+1}{2m+1}\right\rceil\right)^{-1} \sum_{j\in[0,n]} Q_T(T(j)) && (260)\\ =& \left(\left\lceil \frac{n+1}{2m+1}\right\rceil\right)^{-1}, && (261) \end{aligned}$$

where

- the inequality in (259) is from that the average probability of $B_m(T(i))$ over $i \in \mathcal{I}_T$ is no less than the minimal probability of $B_m(T(i))$ for $i \in \mathcal{I}_T$;
- the equality in (260) is from that the cardinality of $\mathcal{I}$ defined in (254) is $\lceil \frac{n+1}{2m+1} \rceil$;
- the equality in (261) is from that for any distribution over types $T(j)$ with $j \in [0,n]$, the sum of $Q_T(T(j))$ over $j \in [0,n]$ is 1.

To bound $q^\star$ from below, we construct a distribution Q'_T as

$$Q'_T(T(j)) = \begin{cases} \left(\left\lceil \frac{n+1}{2m+1}\right\rceil\right)^{-1} & j \in I_T \\ 0 & \textit{otherwise}. \end{cases} \tag{262}$$

By (257) for each $i \in [0,n]$, there is a *unique* j satisfying $|i-j| \le m$. Therefore, we bound (253) by

$$\begin{aligned} q^\star \ge& \min_i Q'_T(B_m(T(i))) && (263)\\ =& \min_i Q'_T\Big(\bigcup_{\substack{|i-j|\le m \\ j\in\mathcal{I}_T}} T(j)\Big) && (264)\\ =& \left(\left\lceil \frac{n+1}{2m+1}\right\rceil\right)^{-1}, && (265) \end{aligned}$$

where the equality in (265) holds because for any $i \in [0,n]$, there is only one $j \in \mathcal{I}_T$ satisfying $|i-j| \le m$ such that the union in (264) has exactly one element in it.

Therefore, $q^\star = \left(\left\lceil \frac{n+1}{2m+1}\right\rceil\right)^{-1}$ and the Q'_T defined in (262) achieve the optimum in (253). Thus, we can derive an optimal $Q^\star_{Y^n}$, which assigns the same non-zero probability to only one dataset of each type classes indexed by I_T, i.e., $Q^\star_{Y^n}(y^n) = q^\star$ for one $y^n \in T(j)$ for each $j \in I_T$. Therefore, from (55) we have the corresponding optimal privacy mechanism, which maps all input datasets in one input type class to one feasible output dataset with probability 1. □

Appendix M: Proof of Theorem 11

For the Hamming distortion function on datasets in (70), the feasible ball $B_m(x^n)$ of any dataset $x^n \in \mathcal{X}^n$ is given by

$$B_m(x^n) = \left\{ y^n \in \mathcal{X}^n : d_{\mathrm{H}}(x^n, y^n) \le \frac{m}{n} \right\}. \tag{266}$$

For each $x^n \in \mathcal{X}^n$, the number of datasets having different values at exactly $k > 0$ different positions is $\binom{n}{k}(|\mathcal{X}|-1)^k$. Therefore, the number of elements in its feasible ball $B_m(x^n)$ is

$$|B_m(x^n)| = \sum_{i=0}^{m} \binom{n}{i} (|\mathcal{X}|-1)^i, \tag{267}$$

Note that the cardinality $|B_m(x^n)|$ in (267) of a feasible ball is independent of the input dataset. We denote the cardinality as N_{ball}, i.e., $N_{\text{ball}} \triangleq |B_m(x^n)|$. Due to the symmetric property of the Hamming distortion on datasets in (70), i.e., for any two datasets $x_1^n, x_2^n \in \mathcal{X}^n$, $x_1^n \in B_D(x_2)$ if and only if $x_2 \in B_D(x_1)$, we know that each output dataset is in exactly N_{ball}

different feasible balls (the example in Fig. 6 may help to figure out the above relationships). Therefore,

$$q^\star = \sup_{Q_{Y^n}} \inf_{x^n \in \mathcal{X}^n} Q_{Y^n}\left(B_m(x^n)\right) \tag{268}$$

$$\leq \sup_{Q_{Y^n}} \frac{1}{|\mathcal{X}^n|} \sum_{x^n \in \mathcal{X}^n} Q_{Y^n}\left(B_m(x^n)\right) \tag{269}$$

$$= \sup_{Q_{Y^n}} \frac{1}{|\mathcal{X}^n|} \sum_{x^n \in \mathcal{X}^n} \sum_{y^n \in B_m(x^n)} Q_{Y^n}\left(y^n\right) \tag{270}$$

$$= \sup_{Q_{Y^n}} \frac{1}{|\mathcal{X}|^n} \sum_{\substack{x^n \in \mathcal{X}^n \\ y^n \in B_m(x^n)}} Q_{Y^n}\left(y^n\right) \tag{271}$$

$$= \sup_{Q_{Y^n}} \frac{1}{|\mathcal{X}|^n} \sum_{y^n \in \mathcal{X}^n} N_{\text{ball}} Q_{Y^n}\left(y^n\right) \tag{272}$$

$$= \frac{N_{\text{ball}}}{|\mathcal{X}|^n} \tag{273}$$

where

- the equality in (269) holds if and only if for an arbitrary pair of datasets x_1^n, x_2^n, there is

$$Q_{Y^n}\left(B_D(x_1^n)\right) = Q_{Y^n}\left(B_D(x_2^n)\right), \tag{274}$$

  which can be satisfied by a uniform distribution over $\mathcal{X}^n$, i.e., $Q_{Y^n}^\star = \frac{1}{|\mathcal{X}|^n}$.
- the equality in (272) holds because, for each y^n, the number of sequences x^n where $d_H(x^n, y^n) \leq \frac{m}{n}$ is exactly N_{ball}.

□

Acknowledgment

The authors would like to thank Dr. Mario Alberto Diaz Torres and Prof. Vincent Y. F. Tan for many useful discussions.

References

[1] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644–654, 1976.

[2] T. Elgamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," *IEEE Transactions on Information Theory*, vol. 31, no. 4, pp. 469–472, 1985.

[3] R. D. Prisco and A. D. Santis, "On the relation of random grid and deterministic visual cryptography," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 4, pp. 653–665, 2014.

[4] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," *IEEE Transactions on Information Theory*, vol. 24, no. 4, pp. 451–456, 1978.

[5] N. Bhargav, S. L. Cotton, and D. E. Simmons, "Secrecy capacity analysis over κ-μ fading channels: Theory and applications," *IEEE Transactions on Communications*, vol. 64, no. 7, pp. 3011–3024, 2016.

[6] B. Dai, L. Yu, and Z. Ma, "Relay broadcast channel with confidential messages," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 410–425, 2016.

[7] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *IEEE Symposium on Security and Privacy*, 2008.

[8] T. Ristenpart, E. Tromer, H. Shacham, and S. Savage, "Hey, you, get off of my cloud: Exploring information leakage in third-party compute clouds," in *16th ACM Conference on Computer and Communications Security*, 2009, pp. 199–212.

[9] D. Shah and T. Zaman, "Rumors in a network: Who's the culprit?" *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5163–5181, 2011.

[10] G. Liang, W. He, C. Xu, L. Chen, and J. Zeng, "Rumor identification in microblogging systems based on users' behavior," *IEEE Transactions on Computational Social Systems*, vol. 2, no. 3, pp. 99–108, 2015.

[11] A. Ghassami, X. Gong, and N. Kiyavash, "Capacity limit of queueing timing channel in shared FCFS schedulers," in *IEEE International Symposium on Information Theory*, 2015, pp. 789–793.

[12] A. K. Biswas, "Efficient timing channel protection for hybrid (packet/circuit-switched) network-on-chip," *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 5, pp. 1044–1057, 2018.

[13] C. Dwork, "Differential privacy," in *33rd International Colloquium on Automata, Languages and Programming*, Venice, Italy, Jul. 2006.

[14] C. Dwork, "Differential privacy: A survey of results," in *Theory and Applications of Models of Computation: Lecture Notes in Computer Science*. New York:Springer, Apr. 2008.

[15] C. C. Aggarwal, "On k-anonymity and the curse of dimensionality," in *31st International Conference on Very Large Data Bases*. ACM, 2005, pp. 901–909.

[16] D. Rebollo-Monedero, J. Forne, and J. Domingo-Ferrer, "From t-closeness-like privacy to postrandomization via information theory," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 11, pp. 1623–1636, Nov. 2010.

[17] G. Aggarwal, T. Feder, K. Kenthapadi, S. Khuller, R. Panigraphy, D. Thomas, and A. Zhu, "Achieving anonymity via clustering," in *Symp. Principles of Database Systems*, Dallas, TX, Jun. 2006.

[18] F. du Pin Calmon and N. Fawaz, "Privacy against statistical inference," in *50th Annual Allerton Conference onCommunication, Control, and Computing*, 2012.

[19] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *IEEE Trans. on Inform. For. and Sec.*, vol. 8, no. 6, pp. 838–852, 2013.

[20] L. Sankar, S. K. Kar, R. Tandon, and H. V. Poor, "Competitive privacy in the smart grid: An information-theoretic approach," in *Smart Grid Communications*, Brusells, Belgium, Oct. 2011.

[21] S. Asoodeh, F. Alajaji, and T. Linder, "On maximal correlation, mutual information and data privacy," in *IEEE 14th Canadian Workshop on Information Theory*, 2015.

[22] W. Wang, L. Ying, and J. Zhang, "On the relation between identifiability, differential privacy, and mutual-information privacy," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5018–5029, 2016.

[23] J. Liao, L. Sankar, V. Y. F. Tan, and F. P. Calmon, "Hypothesis testing under mutual information privacy constraints in the high privacy regime," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 4, pp. 1058–1071, 2018.

[24] S. Li, A. Khisti, and A. Mahajan, "Information-theoretic privacy for smart metering systems with a rechargeable battery," *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3679– 3695, 2018.

[25] B. Rassouli and D. Gündüz, "Optimal utility-privacy trade-off with the total variation distance as the privacy measure," in *arXiv:1801.02505v1 [cs.IT]*, 2018.

[26] I. Mironov, "Rényi differential privacy," in *IEEE 30th Computer Security Foundations Symposium*, 2017, pp. 263–275.

[27] S. Asoodeh, M. Diaz, F. Alajaji, and T. Linder, "Privacy-aware guessing efficiency," in *IEEE International Symposium on Information Theory*, 2017, pp. 754–758.

[28] I. Issa, A. B. Wagner, and S. Kamath, "An operational approach to information leakage," *arXiv:1807.07878*, 2018.

[29] J. L. Massey, "Guessing and entropy," in *IEEE International Symposium on Information Theory*, 1994, p. 204.

[30] D. Malone and W. G. Sullivan, "Guesswork and entropy," *IEEE Transactions on Information Theory*, vol. 50, no. 3, pp. 525–526, 2004.

[31] M. M. Christiansen and K. R. Duffy, "Guesswork, large deviations, and shannon entropy," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 796–802, 2013.

[32] N. Merhav and M. Feder, "Universal prediction," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2124–2147, Oct 1998.

[33] X. Nguyen, M. J. Wainwright, and M. I. Jordan, "On surrogate loss functions and f-divergences," *The Annals of Statistics*, vol. 37, no. 2, pp. 876–904, 2009.

[34] T. A. Courtade and R. D. Wesel, "Multiterminal source coding with an entropy-based distortion measure," in *IEEE International Symposium on Information Theory*, July 2011, pp. 2040–2044.

[35] P. L. Bartlett, M. I. Jordan, and J. D. McAuliffe, "Convexity, classification, and risk bounds," *Journal of the American Statistical Association*, vol. 101, no. 473, pp. 138–156, 2006.

[36] J. Liao, L. Sankar, O. Kosut, and F. P. Calmon, "Robustness of maximal α-leakage to side information," *arXiv:1901.07105 [cs.IT]*, 2019.

[37] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *IEEE 54th Annual Symposium on Foundations of Computer Science*, 2013.

[38] Q. Geng, P. Kairouz, S. Oh, and P. Viswanath, "The staircase mechanism in differential privacy," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 7, pp. 1176–1184, 2015.

[39] K. Kalantari, L. Sankar, and A. D. Sarwate, "Robust privacy-utility tradeoffs under differential privacy and hamming distortion," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2816–2830, 2018.

[40] S. Asoodeh, F. Alajaji, and T. Linder, "Privacy-aware MMSE estimation," *IEEE International Symposium on Information Theory*, pp. 1989–1993, 2016.

[41] J. Liao, L. Sankar, V. Y. Tan, and F. P. Calmon, "Hypothesis testing under maximal leakage privacy constraints," in *IEEE International Symposium on Information Theory*, 2017.

[42] E. Tuncel, P. Koulgi, S. Regunathan, and K. Rose, "Zero-error source coding with maximum distortion criterion," in *Data Compression Conference*, 2002, pp. 92–101.

[43] S. Verdú, "α-mutual information," in *IEEE Information Theory and Applications Workshop*, 2015.

[44] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," in *Colloquia mathematica Societatis János Bolyai*, Kestheley, Hungary, 1975, pp. 41–52.

[45] R. Sibson, "Information radius," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 14, no. 2, pp. 149–160, 1969.

[46] A. Rényi, "On measures of entropy and information," in *4th Berkeley Symposium on Mathematical Statistics and Probability*. The Regents of the University of California, 1961, pp. 547–561.

[47] T. Van Erven and P. Harremos, "Rényi divergence and kullback-leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, 2014.

[48] S.-W. Ho and S. Verdú, "Convexity/concavity of Rényi entropy and α-mutual information," in *IEEE International Symposium on Information Theory*, 2015.

[49] P. L. Bartlett, M. I. Jordan, and J. D. McAuliffe, "Convexity, classification, and risk bounds," *Journal of the American Statistical Association*, vol. 101, no. 473, pp. 138–156, 2006.

[50] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 740–761, Jan 2014.

[51] F. Liese and I. Vajda, "On divergences and informations in statistics and information theory," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4394–4412, Oct 2006.

[52] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2014.

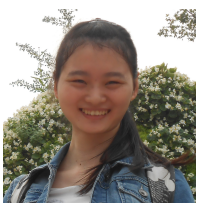
[53] S. Fehr and S. Berens, "On the conditional Rényi entropy," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6801–6810, 2014.

[54] D. Kifer and B.-R. Lin, "An axiomatic view of statistical privacy and utility," *Journal of Privacy and Confidentiality*, vol. 4, no. 1, pp. 5–49, 2012.

[55] Y. Wang, Y. O. Basciftci, and P. Ishwar, "Privacy-utility tradeoffs under constrained data release mechanisms," *arXiv:1710.09295v1 [cs.IT]*, 2017.

[56] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.

[57] J. Liao, L. Sankar, F. P. Calmon, and V. Y. F. Tan, "Hypothesis testing under maximal leakage privacy constraints," in *IEEE International Symposium on Information Theory*, 2017, pp. 779–783.



Jiachun Liao (S'16) received the B.Eng. in communication engineering and M.Eng. degrees in communication and information system from Beijing Jiaotong University, in 2012 and 2015, respectively. She is currently pursuing the Ph.D. degree in the School of Electrical, Computer, and Energy Engineering at Arizona State University. Her research interests includes wireless communications, information privacy and fairness in machine learning.

Oliver Kosut (S'06–M'10) received B.S. degrees in electrical engineering and mathematics from the Massachusetts Institute of Technology, Cambridge, MA, USA in 2004 and the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY, USA in 2010.

Since 2012, he has been a faculty member in the School of Electrical, Computer and Energy Engineering at Arizona State University, Tempe, AZ, USA, where he is an Associate Professor. Previously, he was a Postdoctoral Research Associate in the Laboratory for Information and Decision Systems at MIT from 2010 to 2012. His research interests include information theory, cybersecurity, and power systems. Prof. Kosut received the NSF CAREER award in 2015.

Lalitha Sankar (S'02-M'07-SM'15) received the B.Tech. degree from the Indian Institute of Technology, Bombay, the M.S. degree from the University of Maryland, and the Ph.D. degree from Rutgers University. She is currently an Associate Professor in the School of Electrical, Computer, and Energy Engineering at Arizona State University. Prior to this, she was an Associate Research Scholar at Princeton University and a recipient of a three year Science and Technology Teaching postdoctoral fellowship from the Council on Science and Technology at Princeton University. Sankar has also worked as a Senior Member of Technical Staff at AT&T Shannon Labs and Polaroid Engineering R&D Labs.

Her research interests include applying information sciences to study data privacy as well as cybersecurity and resilience in critical infrastructure networks. For her doctoral work, she received the 2007-2008 Electrical Engineering Academic Achievement Award from Rutgers University. She received the IEEE Globecom 2011 Best Paper Award for her work on privacy of side-information in multi-user data systems, and the National Science Foundation CAREER Award in 2014.

Flavio du Pin Calmon is an Assistant Professor of Electrical Engineering at Harvard's John A. Paulson School of Engineering and Applied Sciences. Before joining Harvard, he was the inaugural data science for social good post-doctoral fellow at IBM Research in Yorktown Heights, New York. He received his Ph.D. in Electrical Engineering and Computer Science at MIT. His main research interests are information theory, inference, and statistics, with applications to fairness, privacy, machine learning, and communications engineering. Prof. Calmon has received the NSF CAREER Award in 2019, the Google Research Faculty Award, the IBM Open Collaborative Research Award, and Harvard's Lemann Brazil Research Fund Award.