Information-Theoretic Privacy with General Distortion Constraints

Kousha Kalantari Student Member, IEEE, Oliver Kosut Member, IEEE, Lalitha Sankar Senior Member, IEEE

Abstract

The privacy-utility tradeoff problem is formulated as determining the privacy mechanism (random mapping) that minimizes the mutual information (a metric for privacy leakage) between the private features of the original dataset and a released version. The minimization is studied with two types of constraints on the distortion between the public features and the released version of the dataset: (i) subject to a constraint on the expected value of a cost function $f$ applied to the distortion, and (ii) subject to bounding the complementary CDF of the distortion by a non-increasing function $g$. The first scenario captures various practical cost functions for distorted released data, while the second scenario covers large deviation constraints on utility. The asymptotic optimal leakage is derived in both scenarios. For the distortion cost constraint, it is shown that for convex cost functions there is no asymptotic loss in using stationary memoryless mechanisms. For the complementary CDF bound on distortion, the asymptotic leakage is derived for general mechanisms and shown to be the integral of the single letter leakage function with respect to the Lebesgue measure defined based on the refined bound on distortion. However, it is shown that memoryless mechanisms are generally suboptimal in both cases.

Index Terms

Privacy-utility tradeoff, mutual information leakage, distortion cost function, distortion distribution constraints.

I. INTRODUCTION

Let $(X^n, Y^n)$ be a random data sequence where $X$ and $Y$ represent the public and private sections of the data respectively, and are drawn from an i.i.d. distribution $P_{X,Y}$. Each entry $(X_i, Y_i)$ represents a row of the dataset. We wish to find a privacy mechanism, i.e. a random mapping, that reveals a sequence $\hat{X}^n$ such that (i) statistical information about $X^n$ can be learned from $\hat{X}^n$, and (ii) as little information as possible about private data $Y^n$ should be revealed by $\hat{X}^n$. These two goals are in conflict, since typically $X$ and $Y$ are correlated (especially when $X = Y$). Thus, we wish to characterize the privacy utility tradeoff (PUT) while being careful to choose meaningful utility and privacy metrics.

Manuscript received August 18, 2017; This paper was presented in part at ISIT 2017 [1]. This work is supported in part by the National Science Foundation under grant CCF-1422358.

The authors are with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, Arizona 85287, USA (e-mail: kousha.kalantari@asu.edu, okosut@asu.edu, lalithasankar@asu.edu).
Our focus is on inferential adversaries that can learn the hidden features $Y^n$ from the released dataset $\hat{X}^n$. To this end, we motivate the use of mutual information between the private features $Y^n$ and the revealed version of the dataset $\hat{X}^n$ as a metric for privacy leakage. We do so by first noting that mutual information is a regret function used in many learning applications and quantifies the Kullback-Leibler distance between the prior and posterior knowledge of the inferred data $Y^n$ from the released data $\hat{X}^n$. Furthermore, mutual information is also related to the Fisher information for the asymptotically large datasets, and combined with Fano’s inequality, it serves as a measure of how well an adversary can estimate functions of hidden features.

For the choice of utility metric, the average distortion constraint in the form of $\mathbb{E}[d(X^n, \hat{X}^n)] \leq D$ has been used in many works, where $D$ is a distortion threshold and $d(\cdot, \cdot)$ is a given distortion function between public data and released data. However, this utility metric does not capture all aspects of distortion distribution. One possible step in order to capture more aspects of the distortion distribution, is via the tail probability constraint (or equivalently called excess distortion constraint). This has been of much interest in source coding (see for example [2]–[6]), channel coding (see for example [7]–[9]) and studied in the context of privacy in [10]. For a more detailed survey on finite blocklength approaches see [11].

However, even the tail probability constraint does not capture the full spectrum of possibilities on applying bounds on distortion distribution. In this paper, we generalize the tail probability constraint in two ways:

- A bound $t$ on the average distortion cost, where the distortion cost is a non-decreasing function $f$ applied on a separable distortion measure $d$ between $X^n$ and $\hat{X}^n$. The resulting PUT is given by

$$\min_{P_{\hat{X}^n|X^n,Y^n}} \frac{1}{n} I(Y^n; \hat{X}^n)$$
subject to
$$\mathbb{E}[f(d(X^n, \hat{X}^n))] \leq t,$$

(1)

- A non-increasing function $g$ to bound the complementary CDF of the distortion measure $d$ between $X^n$ and $\hat{X}^n$. The resulting PUT is given by

$$\min_{P_{\hat{X}^n|X^n,Y^n}} \frac{1}{n} I(Y^n; \hat{X}^n)$$
subject to
$$\mathbb{P}[d(X^n, \hat{X}^n) > D] \leq g(D), \forall D.$$

(2)

The cost constraint in (1) imposes increasing penalties on higher levels of distortion in general, and reduces to a tail probability constraint when $f(D) = 1(D > D_0)$, for some constant $D_0$. The distortion distribution bound in (2) allows arbitrarily fine-tuned bounds on the complementary CDF of the distortion, and reduces to a tail probability constraint when $g(D) = 1 - (1 - \epsilon)1(D \geq D_0)$, for some constant $D_0$. Note that these two types of constraint are not equivalent in general and can capture different requirements on the distortion distribution.

A. Contributions

A privacy mechanism could be applied to a dataset as a whole, or to each individual entry of the dataset independently. We label the mechanisms for the two approaches as general and memoryless mechanisms. In this paper:

- We derive precise expressions for the asymptotic leakage distortion-cost function tradeoff in (1). For memoryless mechanisms, it is equal to the single letter leakage function evaluated at the inverse of the cost function applied
to the cost threshold $t$, and for general mechanisms, it is the lower convex envelope of the leakage tradeoff curve under memoryless mechanisms.

- We also give the exact formulation of the asymptotic leakage in (2) for memoryless and general mechanisms.
  For memoryless mechanisms, it is equal to the single letter leakage function evaluated at the largest distortion value that $g(.)$ is equal to 1. For general mechanisms, it is the integral of single letter leakage function with respect to the Lebesgue measure defined based on the constraint function $g$.

- In both cases, the optimal general mechanisms are mixtures of memoryless mechanisms.

The formulations in (1) and (2) include the dependence on both the public and private aspects of the dataset. In cases where the private data is not directly available, but the statistics are known, the private ($Y^n$), public ($X^n$), and revealed data ($\hat{X}^n$) form a Markov chain $Y^n \rightarrow X^n \rightarrow \hat{X}^n$. In this paper, we focus on the general case with both public and private data being available to the mechanism, but the results here generalize in a straightforward manner to the case when private data is not available.

B. Related Work

An alternative approach to more general distortion constraints is considered in [6] and referred to as $\tilde{f}$-separable distortion measures. In [6], a multi-letter distortion measure $\tilde{d}(\cdot, \cdot)$ is defined as $\tilde{f}$-separable if

$$\tilde{d}(x^n, \hat{x}^n) = \tilde{f}^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} \tilde{f}(\tilde{d}(x_i, \hat{x}_i)) \right),$$

for an increasing function $\tilde{f}$. The distortion cost constraints that we consider are more general in the sense that our notion of cost function $f$ applied to the distortion measure $d(\cdot, \cdot)$ covers a broader class of distortion constraints than an average bound on $\tilde{f}$-separable distortion measures studied in [6]. Specifically, the average constraint on an $\tilde{f}$-separable distortion measure has the form

$$\mathbb{E} \left[ \tilde{f}^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} \tilde{f}(\tilde{d}(x_i, \hat{x}_i)) \right) \right] \leq D,$$

which clearly is a specific case for our formulation in (1) that results from choosing $f = \tilde{f}^{-1}$ and $d(x, \hat{x}) = \tilde{f}(\tilde{d}(x, \hat{x}))$, such that $d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i)$. Moreover, we allow for non-decreasing functions $f$, which means that $\tilde{f}$ does not have to be strictly increasing. We also note that our focus is on privacy rather than source coding.

In the context of privacy, the privacy utility tradeoff with distinct $X$ and $Y$ is studied in [12] and more extensively in [13], but the utility metric is only restricted to identity cost functions, i.e. $f(D) = D$. Generalizing this to the excess distortion constraint was considered by [10]. In [10], we also differentiated between explicit availability or unavailability of the private data $Y$ to the privacy mechanism. Information theoretic approaches to privacy that are agnostic to the length of the dataset are considered in [14]–[16].

In [10], we also allow the mechanisms to be either memoryless (also referred to as local privacy) or general. This approach has also been considered in the context of differential privacy (DP) (see for example [17]–[21]).

1We have changed their notation from $f$-separable to $\tilde{f}$-separable, in order to avoid confusion with our notation.
the information theoretic context, it is useful to understand how memoryless mechanisms behave for more general distortion constraints as considered here. Furthermore, even less is known about how general mechanisms behave and that is what this paper aims to do.

In this paper, we first setup the problem formulation in Section III. Then, in Section III we present our main results for the asymptotic leakage for general and memoryless mechanisms, under the average distortion cost and complementary CDF bounds on distortion. Finally, we provide all the proofs in Sections V.

C. Notation

Throughout this paper we use $D$ as the distortion value, and $d(\cdot, \cdot)$ to indicate the distortion function used for measuring utility. We also use $D_{KL}(\cdot || \cdot)$ for the KL-divergence between two distributions. The mutual information between two variables $X$ and $Y$ is denoted by $I(X; Y)$ and the base for all the logarithm and exponential functions are the same, but can be any numerical value. We denote binary entropy by $H_b(\cdot)$, and use $E_P[\cdot]$ for expectation with respect to distribution $P$, where the subscript $P$ is dropped when it is clear from context. We denote random variables with capital letters, and their corresponding alphabet set by calligraphic letters. The lower convex envelope of a function $r(\cdot)$ for any point $t$ in its domain is given by

$$r^{**}(t) \triangleq \sup \left\{ s(t) \middle| \begin{array}{l} s \text{ is convex}, \cr s(x) \leq r(x), \forall x \in \text{Dom } r \end{array} \right\}. \tag{5}$$

II. Problem Definition and Preliminaries

Let the source data $(X^n, Y^n)$ be a dataset of $n$ independently and identically distributed (i.i.d.) random variables, where $(X_i, Y_i) \sim P_{X,Y}$, for all $i = 1, \ldots, n$. The revealed data is an $n$-length sequence $\hat{X}^n$ drawn from the alphabet $\hat{X}$, and all the alphabet sets $X, Y, \hat{X}$ are assumed to be finite sets. A random mechanism is used to generate the revealed data $\hat{X}^n$ given the source data $(X^n, Y^n)$.

In order to quantify the utility of the revealed data, consider the single letter distortion measure as a function $d: X \times \hat{X} \rightarrow [D_{min}, D_{max}]$. Then, the distortion between $n$-length sequences is given by $d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^n d(x_i, \hat{x}_i)$.

The following definitions represent our main quantities of interest, given by the minimum leakage for a dataset subject to a distortion cost constraint and a complementary CDF bound on distortion. We differentiate between the memoryless and general mechanisms by the superscripts $M$ and $G$, respectively.

**Definition 1 (Information Leakage under a Cost Function):** Given a left-continuous and non-decreasing cost function $f: [D_{min}, D_{max}] \rightarrow [0, \infty)$ and $t > f(D_{min})$, the minimal leakage under an expected distortion cost constraint is defined as follows:

$$L^{(1)}(n, t, f) \triangleq \min_{P_{\hat{X}^n|X^n, Y^n}: \mathbb{E}[f(d(X^n, \hat{X}^n))] \leq t} \frac{1}{n} I(Y^n; \hat{X}^n), \tag{6}$$

and

$$L^{(1)}(t, f) \triangleq \lim_{n \to \infty} L^{(1)}(n, t, f). \tag{7}$$
where the superscript (·) takes values $M$ or $G$. For $L^{(M)}$, the $n$-letter mechanism $P_{X^n|Y^n}$ is restricted to be stationary and memoryless and given by $P_{X^n|Y^n} = (P_X|Y)^n$, while for $L^{(G)}$ it can be any mechanism.

Definition 2 (Information Leakage with Distortion CDF Bound): Given a right-continuous and non-increasing function $g : [D_{\min}, D_{\max}] \rightarrow (0,1]$, the minimal leakage with a cumulative distortion distribution bounded by $g$ is defined as follows:

$$L^{(g)}(n) \triangleq \min_{P_{X^n|Y^n} : \mathbb{E}[d(X^n, Y^n)] \leq g(D)} \frac{1}{n} I(Y^n; \hat{X}^n),$$

(8)

and

$$L^{(g)}(g) \triangleq \lim_{n \to \infty} L^{(g)}(n,g),$$

(9)

where the superscript (·) takes values $M$ or $G$. For $L^{(M)}$, the $n$-letter mechanism $P_{X^n|Y^n}$ is restricted to be stationary and memoryless and given by $P_{X^n|Y^n} = (P_X|Y)^n$, while for $L^{(G)}$ it can be any mechanism.

We now define the optimal single letter information leakage under a constraint on the expected value of the distortion. This is analogous to the single-letter rate-distortion function, and has appeared in earlier works on privacy [13]. As we will show later, this quantity appears as a key element in first-order leakage.

Definition 3 (Single Letter Information Leakage):

$$L(D) \triangleq \min_{P_{X,Y} : \mathbb{E}[d(X,Y)] \leq D} I(Y; \hat{X}).$$

(10)

Note that $L(·)$ is convex, and thus, continuous in $D$.

Remark 1: For $f(D) = D$, and any $n$, the optimization in (8) reduces to (10) for both memoryless and general mechanisms.

We now define functions that will be critical in expressing asymptotic leakage with the expected distortion cost bound under stationary memoryless and general mechanisms.

Definition 4: For any cost function $f$, and a distortion cost threshold $t > f(D_{\min})$, let

$$f^{-1}_{l}(t) \triangleq \sup\{D \in [D_{\min}, D_{\max}] : f(D) < t\},$$

(11)

$$f^{-1}_{u}(t) \triangleq \sup\{D \in [D_{\min}, D_{\max}] : f(D) \leq t\},$$

(12)

and define

$$\mathcal{T}_f \triangleq \{ t : f^{-1}_{l}(t) \neq f^{-1}_{u}(t) \}.$$  

(13)

Consequently, for any $t \notin \mathcal{T}_f$, we have $f^{-1}_{l}(t) = f^{-1}_{u}(t)$, and thus, the inverse function for $f$ can be uniquely determined as

$$f^{-1}(t) \triangleq f^{-1}_{l}(t) = f^{-1}_{u}(t).$$

(14)

III. MAIN RESULTS

A. Distortion Cost Constraint

Theorem 1: Let $t > f(D_{\min})$. If $t \notin \mathcal{T}_f$, then the asymptotic minimum leakage under stationary memoryless mechanisms is given by

$$L^{(M)}(t, f) = (L \circ f^{-1})(t),$$

(15)

August 21, 2017  DRAFT
and for any $t \in T_f$, we have

$$L^{(M)}(t, f) = \min_{P_{X|X,Y}} I(Y; \hat{X}),$$  \hfill (16)

subject to

$$R^u(P_{X,\hat{X}}; c) \triangleq \min_{Q_{X,\hat{X}}: \mathbb{P}[d(X,\hat{X})] \geq c} D_{KL}(Q_{X,\hat{X}} \parallel P_{X,\hat{X}}),$$  \hfill (17)

where for any $P_{X,\hat{X}}$ and constant $c$,

$$R^l(P_{X,\hat{X}}; c) \triangleq \min_{Q_{X,\hat{X}}: \mathbb{P}[d(X,\hat{X})] \leq c} D_{KL}(Q_{X,\hat{X}} \parallel P_{X,\hat{X}}).$$  \hfill (18)

Furthermore, the inequality constraint in (16) reduces to equality if $L(f^{-1}_u(t)) > 0$.

Proof sketch: From the law of large numbers, applying a memoryless mechanism concentrates the distortion around a particular $D$, typically to the expected value, as $n \to \infty$. Therefore, the distortion cost constraint roughly translates to choosing an expected distortion $D$ such that $f(D) \leq t$, or equivalently $f^{-1}(t) \geq D$. If $f^{-1}(t)$ is uniquely determined, then we have the asymptotic leakage in the form of $L(f^{-1}(t))$. Otherwise, our desired $D$ lies somewhere between $f^{-1}_l(t)$ and $f^{-1}_u(t)$. For a more detailed proof, see Section V-A.

Remark 2: If $f(\cdot)$ is strictly increasing, then $T_f = \emptyset$, and $L^{(M)}(t, f)$ is given by (15) for any $t$.

Remark 3: For any $t > f(D_{\min})$, since the closure of the convex hull of epigraphs of $L \circ f^{-1}_l$ and $L \circ f^{-1}_u$ are equal, their lower convex envelopes are equal too. Therefore, $(L \circ f^{-1}_l)^{**}(t) = (L \circ f^{-1}_u)^{**}(t)$, and we refer to this value as $(L \circ f^{-1})^{**}(t)$.

Theorem 2: The asymptotic minimum leakage under general mechanisms is given by

$$L^{(G)}(t, f) = (L \circ f^{-1})^{**}(t).$$  \hfill (19)

Proof sketch: Since $L^{(G)}(t, f)$ is convex in $t$, a convex combination of any two feasible mechanisms is also feasible. Hence, we can always design convex combinations of memoryless mechanisms to achieve the lower convex envelope of $(L \circ f^{-1})(t)$, and therefore $L^{(G)}(t, f) \leq (L \circ f^{-1})^{**}(t)$. Conversely, we show that it is not possible to achieve a smaller leakage. For proof details, we refer the reader to Section V-C.

Remark 4: Note that for $t \geq f(D_{\max})$, we have $L^{(M)}(t, f) = L^{(G)}(t, f) = 0$, where the minimum is achieved by any mechanism with output independent from the input.

Remark 5: If $f$ is convex, then $(L \circ f^{-1})^{**}(t) = L(f^{-1}(t))$. Therefore, from Theorem I we have

$$L^{(G)}(t, f) = L^{(M)}(t, f) = L(f^{-1}(t)).$$  \hfill (20)

Remark 6: Note that if $L(f^{-1}(t))$ is not equal to its lower convex envelope for some $t$, then the optimal mechanism is formed by a convex combination of the optimal memoryless mechanisms for distortion costs $t_1$ and $t_2$, where $t_1$ is the largest threshold smaller than $t$ and $t_2$ is the smallest threshold larger than $t$, such that $L(f^{-1}(\cdot))$ is equal to its lower convex envelope at $t_1$ and $t_2$. 

August 21, 2017
B. Complementary CDF Bound

We now proceed to the result on information leakage with distortion CDF bound. In the following, we give closed form results for the asymptotic information leakage with the distortion CDF bounded by a function \( g \).

**Theorem 3:** If \( g \) is a non-increasing right-continuous function, then the asymptotic information leakage for memoryless mechanisms under a distortion CDF bound is given by

\[
L^{(M)}(g) = L(D_g),
\]

(21)

where \( D_g \equiv \inf\{ D \in [D_{\min}, D_{\max}] : g(D) < 1 \} \).

**Proof:** Suppose \( D_g > D_{\min} \). Then, for any fixed \( \delta > 0 \) and \( n \), choose \( P_{\hat{X}^n|X^n,Y^n} = \left( P_{X^n|X,Y}^{(n)} \right)^n \), where \( P_{X^n|X,Y}^{(n)} \) is the optimal single letter mechanism achieving \( L(D_g - \delta) \). Since \( g \) is bounded away from zero and \( \mathbb{P}[d(X^n, \hat{X}^n) > D_g] \) goes to zero as \( n \) goes to infinity, the distortion constraint \( \mathbb{P}[d(X^n, \hat{X}^n) > D] \leq g(D) \) is satisfied for all \( D \) for sufficiently large \( n \). Then, as \( \delta \to 0 \), continuity of \( L(\cdot) \) implies \( L(D_g) \) is achievable.

Conversely, according to the law of large numbers, the distortion \( d(X^n, \hat{X}^n) \) concentrates around its expected value as \( n \) goes to infinity. In other words, we have \( \mathbb{P}[d(X^n, \hat{X}^n) > D] \to 1 \), if \( D < \mathbb{E}[d(X^n, \hat{X}^n)] \). This, in turn, implies that for any \( D \) such that \( g(D) < 1 \), we must have \( \mathbb{E}[d(X^n, \hat{X}^n)] < D \). Therefore, a feasible memoryless mechanism has to satisfy \( \mathbb{E}[d(X^n, \hat{X}^n)] \leq D_g \).

Finally, for \( D_g = D_{\min} \), we have to satisfy \( \mathbb{P}[d(X^n, \hat{X}^n) = D_{\min}] = 1 \). Note that in this case, the constraint for \( L(D_g) \), i.e. \( \mathbb{E}[d(X^n, \hat{X}^n)] \leq D_g \), is also equivalent to \( \mathbb{P}[d(X^n, \hat{X}^n) = D_{\min}] = 1 \). Therefore, the set of feasible memoryless mechanisms for \( L^{(M)}(g) \) is equal to those for \( L(D_g) \), and thus, \( L^{(M)}(g) = L(D_g) \).

**Theorem 4:** If \( g \) is a non-increasing right-continuous function, and the single letter leakage function \( L(\cdot) \) is bounded on \([D_{\min}, D_{\max}]\), then

\[
L^{(G)}(g) = \int_{D_{\min}}^{D_{\max}} L(D) d(g(D)),
\]

(22)

where the integral is a Lebesgue–Stieltjes integral of the single letter leakage function \( L(\cdot) \) with respect to the Lebesgue–Stieltjes measure associated with the constraint function \( g \).

**Proof sketch:** We first prove this result for simple constraint functions \( g \), which are in the form of a finite sum of step functions. Then, we show that any non-increasing right-continuous constraint function \( g \) can be upper and lower bounded by such simple functions, and therefore, the corresponding leakage can be upper and lower bounded by that of the simple functions. For a more detailed proof, see Section V-D.

**Remark 7:** An alternative way of describing the result in Theorem 4 is that the asymptotically optimal mechanism behaves as if it first chooses a random \( D \) drawn from a distribution with a complementary CDF exactly equal to \( g(\cdot) \), and then applies the single letter optimal mechanism achieving the single letter optimal leakage \( L(D) \) in a stationary and memoryless fashion. Thus, averaging over the random choice of \( D \), the resulting leakage is given as the integral in (22).

C. Auxiliary Result

We now present a result characterizing the asymptotic optimal privacy leakage subject to multiple excess probability constraints. This can be seen as a special case of complementary CDF bound in which the \( g \) function is a
simple function, i.e. it takes finitely many values. The following results will also be used in the proof of Theorem 2.

For vectors $D = (D_1, D_2, \ldots, D_k)$ and $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_k)$, where $D_{\min} \leq D_1 < \cdots < D_k \leq D_{\max}$ and $1 \geq \epsilon_1 > \cdots > \epsilon_k > 0$, a simple function $g_{\epsilon, D}$ is illustrated in Fig. 1 and formally defined as

$$g_{\epsilon, D}(D) \triangleq \begin{cases} 1, & D_{\min} \leq D < D_1, \\ \epsilon_i, & D_i \leq D < D_{i+1}, i = 1, \ldots, k - 1, \\ \epsilon_k, & D_k \leq D \leq D_{\max}. \end{cases} \quad (23)$$

One can verify that for a constraint function of this form, the minimization in (9) is equivalent to the information leakage with multiple excess distortion constraints, defined as follows.

**Definition 5 (Information Leakage with Multiple Excess Probability Constraints):** Given a distortion vector $D = (D_1, D_2, \ldots, D_k)$ and a tail probability vector $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_k)$, where $D_{\min} \leq D_1 < \cdots < D_k \leq D_{\max}$ and $1 \geq \epsilon_1 > \cdots > \epsilon_k > 0$, the minimal leakage with multiple excess distortion constraints is defined as

$$L^{(G)}(n, D, \epsilon) \triangleq \min_{P_{\hat{X}^n|X^n, Y^n} : \Pr[d(X^n, \hat{X}^n) > D_i] \leq \epsilon_i, \forall 1 \leq i \leq k} \frac{1}{n} I(Y^n; \hat{X}^n), \quad (24)$$

where the $n$-letter mechanisms in (6) are not constrained to be memoryless or stationary, and

$$L^{(G)}(D, \epsilon) \triangleq \lim_{n \to \infty} L^{(G)}(n, D, \epsilon). \quad (25)$$

In the following lemma, we provide the asymptotic optimal leakage under general mechanisms for the class of distortion CDF bound functions defined in Definition 5.

**Lemma 1:**

$$L^{(G)}(D, \epsilon) = \sum_{i=1}^{k} (\epsilon_{i-1} - \epsilon_i) L(D_i) = \int_{D_{\min}}^{D_{\max}} L(D) d(g_{\epsilon, D}(D)), \quad (26)$$

where $\epsilon_0 = 1$. In particular, we have

$$L^{(G)}(n, D, \epsilon) = \sum_{i=1}^{k} (\epsilon_{i-1} - \epsilon_i) L(D_i) + \theta(k, n), \quad (27)$$
where
\[-\frac{\log(k+1)}{n} \leq \theta(k,n) \leq O\left(\sqrt{\frac{\log n}{n}}\right). \tag{28}\]

Proof sketch: The proof hinges on choosing a combination of memoryless mechanisms, each of them being the single letter optimal mechanism for a separate $D_i$ applied in a stationary and memoryless fashion. The weights of this combination will be chosen such that all the excess distortion probabilities are met. For a detailed proof see section V.B.

IV. ILLUSTRATION OF RESULTS

In this section, we first examine the generic cases of single and double step $f$ and $g$ functions. Then, we consider a doubly symmetric binary source and derive its corresponding single letter leakage function. Finally, we use the single letter leakage function to find the asymptotically optimal leakage under specific examples of the average distortion cost constraint and complementary CDF bound.

A. Distortion Cost Function

Example 1: $f(D) = 1(D > D_0)$ as shown in Fig. 2. In this case, $T_f = \{1\}$, and we have

\[f^{-1}_u(t) = \begin{cases} D_0, & t < 1, \\ D_{\text{max}}, & t \geq 1, \end{cases}\] \tag{29}

\[f^{-1}_l(t) = \begin{cases} D_0, & t \leq 1, \\ D_{\text{max}}, & t > 1. \end{cases}\] \tag{30}

Therefore, according to Theorem 1 for stationary memoryless mechanisms we have

\[L^{(M)}(t,f) = \begin{cases} L(D_0), & 0 < t < 1, \\ 0, & t \geq 1, \end{cases}\] \tag{31}

and for general mechanisms, according to Theorem 2 we have

\[L^{(G)}(t,f) = \begin{cases} (1-t)L(D_0), & 0 \leq t < 1 \\ 0, & t \geq 1. \end{cases}\] \tag{32}

Fig. 2. The single step cost function $f(D) = 1(D > D_0)$. 

August 21, 2017 DRAFT
This exactly matches our earlier results in [10] and for the special case of $X = Y$ simplifies to the result in [3]. The leakages $L^G(t, f)$ and $L^M(t, f)$ are depicted in Fig. 3. Note that for $t = 1$, we have $L^G(t, f) = L^M(t, f) = 0$ due to Remark 4.

Example 2: $f(D) = a_1 1(D > D_1) + a_2 1(D > D_2)$, $D_1 < D_2$ as shown in Fig. 4. In this case, $\mathcal{T}_f = \{a_1, a_1 + a_2\}$, and we have

$$f_u^{-1}(t) = \begin{cases} D_1, & t < a_1, \\ D_2, & a_1 \leq t < a_1 + a_2, \\ D_{\text{max}}, & t \geq a_1 + a_2, \end{cases}$$

(33)

$$f_l^{-1}(t) = \begin{cases} D_1, & t \leq a_1, \\ D_2, & a_1 < t \leq a_1 + a_2, \\ D_{\text{max}}, & t > a_1 + a_2. \end{cases}$$

(34)

Hence, according to Theorem 1 for stationary memoryless mechanisms we have

$$L^M(t, f) = \begin{cases} L(D_1), & t < a_1, \\ L(D_2), & a_1 < t < a_1 + a_2, \\ 0, & a_1 + a_2 \leq t. \end{cases}$$

(35)

Note that for $t = a_1$, the exact value for $L^M(t, f)$ is derived by (16), and for $t = a_1 + a_2$, we have $L^M(t, f) = 0$ due to Remark 4. From Theorem 2 we know that $L^G(t, f)$ is the lower convex envelope of $L^M(t, f)$. If $a_2 L(D_1) \geq (a_1 + a_2) L(D_2)$, then it is given by

$$L^G(t, f) = \begin{cases} L(D_2) + (1 - \frac{t}{a_1}) L(D_1), & t \leq a_1, \\ (1 - \frac{t - a_2}{a_2}) L(D_2), & a_1 \leq t \leq a_1 + a_2, \\ 0, & a_1 + a_2 \leq t, \end{cases}$$

(36)

and otherwise,

$$L^G(t, f) = \begin{cases} (1 - \frac{t}{a_1 + a_2}) L(D_1), & t \leq a_1 + a_2, \\ 0, & a_1 + a_2 \leq t. \end{cases}$$

(37)

These two cases together with their corresponding $L^M(t, f)$ are shown in Figs. 5 and 6, respectively.
Fig. 4. The double step cost function \( f(D) = a_1 \mathbb{1}(D > D_1) + a_2 \mathbb{1}(D > D_2) \), \( D_1 < D_2 \).

Fig. 5. \( L^M(t, f) \) and \( L^G(t, f) \) for \( f(D) = a_1 \mathbb{1}(D > D_1) + a_2 \mathbb{1}(D > D_2) \), if \( a_2 L(D_1) \geq (a_1 + a_2) L(D_2) \).

Fig. 6. \( L^M(t, f) \) and \( L^G(t, f) \) for \( f(D) = a_1 \mathbb{1}(D > D_1) + a_2 \mathbb{1}(D > D_2) \), if \( a_2 L(D_1) < (a_1 + a_2) L(D_2) \).

B. Distortion CDF Constraints

We now proceed to complementary CDF bounds on distortion. First, we consider a single step function \( g \) (hard tail probability constraint), and then generalize to a sum of two step functions.

Example 3: \( g(D) = 1 - (1 - \epsilon) \mathbb{1}(D \geq D_0) \) as shown in Fig. 7, where \( D_{\text{min}} < D_0 < D_{\text{max}} \). For stationary memoryless mechanisms we have

\[
L^M(g) = L(D_0),
\]

while for the general mechanisms, we have

\[
L^G(g) = \int_{D_{\text{min}}}^{D_{\text{max}}} L(D) d(g(D)) = (1 - \epsilon)L(D_0).
\]

Note that this is equivalent to Example 1. Therefore, (38) and (39) verify the results in [3] and [10], wherein the tail probability constraint is used as a utility metric.
Example 4: \( g(D) = 1(D < D_1) + \epsilon_1 1(D_1 \leq D < D_2) + \epsilon_2 1(D_2 \leq D) \) as shown in Fig. 8. For stationary memoryless mechanisms we have
\[
L^{(M)}(g) = L(D_1),
\]
while for the general mechanisms, we have
\[
L^{(G)}(g) = (1 - \epsilon_1)L(D_1) + (\epsilon_1 - \epsilon_2)L(D_2).
\]

C. Doubly Symmetric Binary Source (DSBS)

We now consider a doubly symmetric source with parameter \( q \) as depicted in Fig. 9 with Hamming distortion, i.e. \( d(x, \hat{x}) = 1(x \neq \hat{x}) \), as the utility metric. In the following lemma, proved in Section V-E, we derive the single letter leakage function for this source.
Lemma 2: For a doubly symmetric source with $q \leq 0.5$, the single letter leakage function is given by

$$L(D) = \begin{cases} 
1 - H_b(q + D), & D < 0.5 - q, \\
0, & D \geq 0.5 - q.
\end{cases}$$  \hspace{1cm} (42)$$

Remark 8: Due to the inherent symmetry of the problem, for all $q > 0.5$, Lemma 2 holds with $q$ replaced by $1 - q$.

Given the single letter leakage function for a doubly symmetric source, we provide numerical examples for the asymptotically optimal leakages under both distortion cost constraints and complementary CDF bounds.

Example 5: For a doubly symmetric source with parameter $q = 0.1$ and Hamming distortion, consider the cost function $f(D)$ shown in Fig. 10. Then, the corresponding leakage functions $L^{(M)}(t, f)$ and $L^{(G)}(t, f)$ are shown in Fig. 11.

![Fig. 10. The cost function $f(D)$ for Example 5.](image1)

![Fig. 11. Memoryless and general leakage functions $L^{(M)}(t, f)$ and $L^{(G)}(t, f)$ for Example 5.](image2)

We now proceed to an examples that resemble a soft single step complementary CDF bound. We choose functions that are parametrized with a parameter $\lambda$ such that they converge to a hard single step CDF bound as $\lambda \to \infty$. 
Example 6: Consider a doubly symmetric source with parameter $q = 0.1$. Then, for any $\lambda \geq 0$ define

$$g_\lambda(D) = \epsilon + (1 - \epsilon)1(D \leq D_0) + (1 - \epsilon) \left(1 - \frac{1(D \leq D_0)}{2}\right)e^{-\lambda|D-D_0|}. \quad (43)$$

In Fig. 12 this function is plotted for $D_0 = 0.2$, $\epsilon = 0.1$, and four different values of $\lambda$. Note that in Fig. 13 the value of $L^G(g_\lambda)$ converges to the asymptotic value of $(1 - \epsilon)L(D_0)$ as $\lambda \to \infty$, and $L^G(g_\lambda)$ is non-monotonic in $\lambda$.

![Graph of $g(d)$](image1)

**Fig. 12.** $g(d)$ as described in Example 6 for $D_0 = 0.2$ and $\epsilon = 0.1$, parametrized by $\lambda$.

![Graph of $L(G)(g)$](image2)

**Fig. 13.** $L(G)(g)$ for the $g$ function given in Example 6.

V. PROOFS

Before proving our main results, we first review Hoeffding’s inequality, a version of Chernoff bound used for bounded random variables.

Lemma 3 (Hoeffding’s inequality [22, Theorem 2]): Let $X_1, \ldots, X_n$ bounded independent random variables, i.e. $a_i \leq X_i \leq b_i$ for each $1 \leq i \leq n$. We define the empirical mean of these variables by $\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$. Then

$$\mathbb{P}(\bar{X} - \mathbb{E}[\bar{X}] \geq t) \leq \exp \left(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right). \quad (44)$$
where \( t \) is positive, and \( E[X] \) is the expected value of \( X \).

**A. Proof of Theorem**

Assuming a stationary memoryless mechanism, we provide upper and lower bounds on \( \mathbb{E}(f(d(X^n, \hat{X}^n))) \) in terms of \( f(\mathbb{E}[d(X^n, \hat{X}^n)]) \). This in turn allows us to bound \( L^{(M)}(\cdot) \) in terms of \( L(f^{-1}(\cdot)) \). Let \( \delta_n = (D_{\max} - D_{\min}) \sqrt{\log n / n} \). Then, for large enough \( n \) we have

\[
\mathbb{E} \left[ f(d(X^n, \hat{X}^n)) \right] \leq \mathbb{P} \left( d(X^n, \hat{X}^n) \leq \mathbb{E}[d(X^n, \hat{X}^n)] + \delta_n \right) + f \left( \mathbb{E}[d(X^n, \hat{X}^n)] + \delta_n \right)
\]

\[
+ \mathbb{P} \left( d(X^n, \hat{X}^n) > \mathbb{E}[d(X^n, \hat{X}^n)] + \delta_n \right) f(D_{\max})
\]

\[
\leq f(\mathbb{E}[d(X^n, \hat{X}^n)]) + f(D_{\max}) e^{-n \frac{\delta_n^2}{D_{\max} - D_{\min}^2}}
\]

\[
\leq f \left( \mathbb{E}[d(X^n, \hat{X}^n)] + \delta_n \right) + \frac{f(D_{\max})}{n},
\]

where (45c) is due to Lemma 3 and (45d) follows from the definition of \( \delta_n \). If \( \mathbb{E} \left[ d(X^n, \hat{X}^n) \right] \leq f^{-1}(t - \frac{f(D_{\max})}{n} - \delta_n) \), then \( \mathbb{E} \left[ f \left( d(X^n, \hat{X}^n) \right) \right] < t \), and we have

\[
L^{(M)}(n, t, f) \leq L \left( f_{\left( t - \frac{f(D_{\max})}{n} \right) - \delta_n}^{-1} \right).
\]

Since \( f_{\left( t - \frac{f(D_{\max})}{n} \right) - \delta_n}^{-1}(\cdot) \) is left-continuous, and \( L(\cdot) \) is continuous, taking the limit as \( n \to \infty \) gives

\[
L^{(M)}(t, f) \leq L \left( f_{t}^{-1}(t) \right).
\]

With a similar argument and using the negative of the distortion function in Lemma 3 we have

\[
\mathbb{E} \left[ f(d(X^n, \hat{X}^n)) \right] \geq \mathbb{P} \left( d(X^n, \hat{X}^n) \geq \mathbb{E}[d(X^n, \hat{X}^n)] - \delta_n \right) + f \left( \mathbb{E}[d(X^n, \hat{X}^n)] - \delta_n \right)
\]

\[
+ \mathbb{P} \left( d(X^n, \hat{X}^n) < \mathbb{E}[d(X^n, \hat{X}^n)] - \delta_n \right) f(D_{\min})
\]

\[
\geq \left( 1 - \mathbb{P} \left( d(X^n, \hat{X}^n) < \mathbb{E}[d(X^n, \hat{X}^n)] - \delta_n \right) \right) \cdot f(\mathbb{E}[d(X^n, \hat{X}^n)]) - \delta_n
\]

\[
\geq \left( 1 - \frac{1}{n} \right) f \left( \mathbb{E}[d(X^n, \hat{X}^n)] - \delta_n \right),
\]

where (48d) is due to Lemma 3. Therefore, if

\[
\mathbb{E} \left[ f \left( d(X^n, \hat{X}^n) \right) \right] \leq t,
\]

then

\[
\mathbb{E} \left[ d(X^n, \hat{X}^n) \right] \leq f_{\left( t - \frac{1}{n - 1} \right)^{-1}}(t \left( 1 + \frac{1}{n - 1} \right)) + \delta_n,
\]
and we have
\[
L \left( f_u^{-1} \left( t \left( 1 + \frac{1}{n-1} \right) \right) + \delta_n \right) \leq L^{(M)}(n, t, f). \tag{51}
\]
Since \( f_u^{-1}(\cdot) \) is right-continuous, and \( L(\cdot) \) is continuous, taking the limit as \( n \to \infty \) gives
\[
L \left( f_u^{-1}(t) \right) \leq L^{(M)}(t, f). \tag{52}
\]

Recall the definition of \( T_f \) in (12). If \( t \notin T_f \), then \( f_i^{-1}(t) = f_u^{-1}(t) \). Hence, (47) and (52) imply (15). Otherwise, if \( t \in T_f \), fix some \( \delta > 0 \). Then, we bound the expected distortion cost for the function \( f \) under any mechanism \( P_{\hat{X}|X,Y} \). Specifically, as an upper bound we have
\[
\E \left[ f(d(X^n, \hat{X}^n)) \right] \\
\leq \mathbb{P} \left[ d(X^n, \hat{X}^n) \leq f_i^{-1}(t) - \delta \right] f_i^{-1}(t) - \delta \\
+ \mathbb{P} \left[ f_i^{-1}(t) - \delta < d(X^n, \hat{X}^n) < f_u^{-1}(t) \right] t \\
+ \mathbb{P} \left[ f_u^{-1}(t) \leq d(X^n, \hat{X}^n) \right] f(D_{\text{max}}) \\
= t + (f(D_{\text{max}}) - t) \mathbb{P} \left[ f_u^{-1}(t) \leq d(X^n, \hat{X}^n) \right] \\
- (t - f(f_i^{-1}(t) - \delta)) \mathbb{P} \left[ d(X^n, \hat{X}^n) \leq f_i^{-1}(t) - \delta \right] \\
\leq t + (f(D_{\text{max}}) - t) e^{-n(R^u(P_{X,\hat{X}}, f_u^{-1}(t) - \gamma_n))} \\
- (t - f(f_i^{-1}(t) - \delta)) e^{-n(R^l(P_{X,\hat{X}}, f_i^{-1}(t) - \delta) + \gamma_n)}, \tag{53a}
\]
where (57c) is due to Sanov’s theorem [23, Theorem 11.4.1], \( P_{X,\hat{X}}^h(\cdot) \) and \( P_{X,\hat{X}}^l(\cdot) \) are defined in (17) and (18) respectively, and \( \gamma_n = \frac{|X||\hat{X}|\log(n+1)}{n} \). Therefore, if
\[
(f(D_{\text{max}}) - t) e^{-n(R^u(P_{X,\hat{X}}, f_u^{-1}(t) - \gamma_n))} \\
\leq (t - f(f_i^{-1}(t) - \delta)) e^{-n(R^l(P_{X,\hat{X}}, f_i^{-1}(t) - \delta) + \gamma_n)}, \tag{54}
\]
then \( \E[f(d(X^n, \hat{X}^n))] \leq t \). Note that (54) holds for sufficiently large \( n \) if
\[
R^u \left( P_{X,\hat{X}}, f_u^{-1}(t) \right) > R^l \left( P_{X,\hat{X}}, f_i^{-1}(t) - \delta \right). \tag{55}
\]
Therefore, \( L^{(M)}(t, f) \) is upper bounded by
\[
\inf_{P_{X,\hat{X}}} I(Y; \hat{X}) \\
\text{subject to: } R^l(P_{X,\hat{X}}, f_i^{-1}(t) - \delta) < R^u(P_{X,\hat{X}}, f_u^{-1}(t)). \tag{56}
\]
Conversely, we can lower bound the expected distortion cost using a similar argument used in (57). Thus, we have:
\[
\E \left[ f(d(X^n, \hat{X}^n)) \right] \\
\geq \mathbb{P} \left[ d(X^n, \hat{X}^n) \leq f_i^{-1}(t) \right] f(D_{\text{min}}) \\
+ \mathbb{P} \left[ f_i^{-1}(t) < d(X^n, \hat{X}^n) < f_u^{-1}(t) + \delta \right] t \\
+ \mathbb{P} \left[ f_u^{-1}(t) + \delta \leq d(X^n, \hat{X}^n) \right] f(f_u^{-1}(t) + \delta) \tag{57a}
\]
\[ t + \left( f(f^{-1}_u(t) + \delta) - t \right) \mathbb{P} \left[ f^{-1}_u(t) + \delta \leq d(X^n, \hat{T}^n) \right] \\
- (t - f(D_{\min})) \mathbb{P} \left[ d(X^n, \hat{T}^n) \leq f^{-1}_i(t) \right] \geq t + \left( f(f^{-1}_u(t) + \delta) - t \right) e^{-n(R^x(P_{X, \hat{X}}, f^{-1}_u(t) + \delta) + \gamma_n)} \\
- (t - f(D_{\min})) e^{-n(R^x(P_{X, \hat{X}}, f^{-1}_i(t)) - \gamma_n)}, \quad (57b) \]

If a mechanism \( P_{X|X,Y} \) satisfies \( \mathbb{E} \left[ f(d(X^n, \hat{T}^n)) \right] \leq t \), then

\[ (f(f^{-1}_u(t) + \delta) - t) e^{-n(R^x(P_{X, \hat{X}}, f^{-1}_u(t) + \delta) + \gamma_n)} \\
- (t - f(D_{\min})) e^{-n(R^x(P_{X, \hat{X}}, f^{-1}_i(t)) - \gamma_n)} \leq 0. \quad (58) \]

If (58) holds for sufficiently large \( n \), then

\[ R^x(P_{X, \hat{X}}, f^{-1}_u(t) + \delta) \geq R^x(P_{X, \hat{X}}, f^{-1}_i(t)). \quad (59) \]

Hence, \( L^M(t, f) \) is lower bounded by

\[ \inf_{P_{X|X,Y}} I(Y; \hat{X}) \]

subject to: \[ R^x(P_{X, \hat{X}}, f^{-1}_u(t)) \leq R^x(P_{X, \hat{X}}, f^{-1}_i(t)) + \delta. \quad (60) \]

Since \( D_{KL}(P||Q) \) is continuous in \( P \) and \( Q \), taking the limit in (56) and (60) as \( \delta \to 0 \) gives

\[ L^M(t, f) = \min_{P_{X|X,Y}} I(Y; \hat{X}) \]

subject to: \[ R^x(P_{X, \hat{X}}, f^{-1}_u(t)) \leq R^x(P_{X, \hat{X}}, f^{-1}_i(t)). \quad (61) \]

We now prove that if \( L(f^{-1}_u(t)) > 0 \), then there exists an optimal mechanism for (61) that satisfies the constraint with equality. Let \( \hat{P}_{X|X,Y} \) be an optimal mechanism for \( L^M(t, f) \), which satisfies

\[ R^x(\hat{P}_{X, \hat{X}}, f^{-1}_u(t)) \leq R^x(\hat{P}_{X, \hat{X}}, f^{-1}_i(t)) \quad (62) \]

Also, let \( P^*_{X|X,Y} \) be an optimal mechanism for \( L(f^{-1}_u(t)) \). Since \( L(f^{-1}_u(t)) > 0 \), we know that \( \mathbb{E}_{P^*} [d(X^n, \hat{T}^n)] = f^{-1}_u(t) \). Therefore, we have

\[ R^x(P^*_{X, \hat{X}}, f^{-1}_u(t)) \geq R^x(P^*_{X, \hat{X}}, f^{-1}_i(t)) = 0. \quad (63) \]

According to (52), \( P^*_{X|X,Y} \) achieves a lower leakage than that of \( \hat{P}_{X|X,Y} \). Since \( D(P||Q) \) is a continuous function in both \( P \) and \( Q \) and the mutual information is convex in the conditional distribution, there exists a mechanism \( \tilde{P}_{X|X,Y} \) on the line connecting \( \hat{P}_{X|X,Y} \) and \( P^*_{X|X,Y} \) that satisfies

\[ R^x(\tilde{P}_{X, \hat{X}}, f^{-1}_u(t)) = R^x(\hat{P}_{X, \hat{X}}, f^{-1}_u(t)), \quad (64) \]

and achieves a leakage at most equal to that of \( \hat{P}_{X|X,Y} \). Therefore, it suffices to replace the constraint in (61) with equality.
B. Proof of Lemma 7

Achievability: We build a combination of memoryless mechanisms to show achievability. Specifically, we pick the optimal mechanisms for single letter leakage functions evaluated at approximately $D_1, D_2, \ldots, D_k$. The reason for not choosing the exact values of $D_i$ is that we need the optimal single letter mechanism to satisfy a slightly smaller average distortion bound so that a tail probability constraint is guaranteed.

Recall that $P^*(D)$ is the set of optimal single letter mechanisms for $L(D)$. Then, for any $D$ let $P^{*(D)}_{X|X,Y} \in P^*(D)$ and $P^*_{X^n|X^n,Y^n} = \left( P^{*(D)}_{X|X,Y} \right)^n$. Define $\epsilon_0(n) = \epsilon_0 = 1$, and for any $1 \leq i \leq k$ let

$$D_i(n) \equiv D_i - D_{\text{max}} \sqrt{\frac{\log n}{n}},$$

$$\epsilon_i(n) \equiv \frac{\epsilon_i(n)}{1 - \frac{1}{n}}.$$

For the special case where $D_1 = D_{\text{min}}$, let $D_1(n) = D_1$ instead. Note that for sufficiently large $n$ we have $0 \leq D_i(n) \leq D_i$ and $0 \leq \epsilon_i(n) \leq \epsilon_i$, which implies that

$$\epsilon_i(n) = \epsilon_i + O\left( \frac{1}{n} \right).$$

Now let $E$ be a random variable with alphabet set $\{1, \ldots, k+1\}$, where $P(E = i) = \epsilon_{i-1}(n) - \epsilon_i(n)$ for $1 \leq i \leq k$, and $P(E = k+1) = \epsilon_k(n)$. Then, consider the following mechanism:

$$P_{X^n|X^n,Y^n}(\hat{x}^n|x^n, y^n)$$

$$= \begin{cases} P^{*(D)}_{X^n|X^n,Y^n}(\hat{x}^n|x^n, y^n), & \text{if } 1 \leq E \leq k, \\ P^{*(D)}_{X^n|X^n,Y^n}(\hat{x}^n), & \text{if } E = k+1. \end{cases}$$

First, we show that it is feasible, i.e. it satisfies $P(d(X^n, \hat{X}^n) > D_i) \leq \epsilon_i$ for any $1 \leq i \leq k$. Since $D_i(n) \to D_i$, and $D_i$ has a distinct value for each $i$, there exists a $\delta(i) > 0$ and $n_i$ such that $\delta(i) < e^{-\frac{(D_i-D_i(n))^2}{2D_{\text{max}}^2}}$ for $n \geq n_i$.

Therefore, for any $1 \leq i \leq k$ and $n \geq n_i$ we can bound the $i$th error probability by

$$P[d(X^n, \hat{X}^n) > D_i]$$

$$= \epsilon_{k+1}(n) P(d(X^n, \hat{X}^n) > D_i|E = k+1)$$

$$+ \sum_{j=1}^{k} \left( \epsilon_{j-1}(n) - \epsilon_j(n) \right) P(d(X^n, \hat{X}^n) > D_i|E = j)$$

$$\leq \epsilon_i(n) + \sum_{j=1}^{i} \left( \epsilon_{j-1}(n) - \epsilon_j(n) \right) P(d(X^n, \hat{X}^n) > D_i|E = j)$$

$$\leq \epsilon_i(n) + \sum_{j=1}^{i} \left( \epsilon_{j-1}(n) - \epsilon_j(n) \right) e^{-\frac{(D_i-D_i(n))^2}{2D_{\text{max}}^2}}$$

$$\leq \epsilon_i(n) + e^{-\frac{(D_i-D_i(n))^2}{2D_{\text{max}}^2}} \sum_{j=1}^{i-1} (\epsilon_{j-1}(n) - \epsilon_j(n))$$

$$+ \left( \epsilon_{i-1}(n) - \epsilon_i(n) \right) e^{-\frac{(D_i-D_i(n))^2}{2D_{\text{max}}^2}}.$$
where (69b) follows from Lemma 3, (69e) is due to the definition of $D$, Note that in the special case where $L$ random variable $E$

Let $P$ Assume a mechanism $E$

Converse: Assume a mechanism $P_{X^n | X^n, Y^n}$ satisfying the feasibility constraint of (6). Define the indicator random variable $E$ as

$$E = \begin{cases} 
1, & \text{if } d(X^n, \hat{X}^n) \leq D_1, \\
2, & \text{if } D_1 < d(X^n, \hat{X}^n) \leq D_2, \\
\vdots & \\
k + 1, & \text{if } D_k < d(X^n, \hat{X}^n).
\end{cases}$$

(71)

Let $P_{e_i} = P[E \geq i]$ for $i = 1, \cdots, k + 1$ and $\epsilon_0 = 1$. Clearly, for all feasible $P_{\hat{X}^n | X^n, Y^n}$ and $1 \leq i \leq k + 1$, we have $P_{e_i} \leq \epsilon_{i-1}$. Then:

$$I(Y^n; \hat{X}^n)$$

$$= H(Y^n) - H(Y^n | \hat{X}^n)$$

$$\geq H(Y^n) - H(Y^n | \hat{X}^n, E) - I(Y^n; E | \hat{X}^n)$$

$$\geq H(Y^n) - H(Y^n | \hat{X}^n, E) - H(E)$$

$$\geq H(Y^n | E) - H(Y^n | \hat{X}^n, E) - \log(k + 1)$$

(72a)

(72b)

(72c)

(72d)
\[ I(Y^n; \hat{X}^n | E) - \log(k + 1) \]

\[ = \sum_{i=1}^{k+1} (P_{e_i} - P_{e_{i+1}}) I(Y^n; \hat{X}^n | E = i) - \log(k + 1) \]  

\[ \geq \sum_{i=1}^{k} (P_{e_i} - P_{e_{i+1}}) nL(D_i) \]

\[ + P_{e_{k+1}} nL(D_{\text{max}}) - \log(k + 1) \]  

\[ \geq \sum_{i=1}^{k} (\epsilon_{i-1} - \epsilon_{i}) nL(D_i) + \epsilon_k nL(D_{\text{max}}) - \log(k + 1) \]  

\[ \geq \sum_{i=1}^{k} (\epsilon_{i-1} - \epsilon_{i}) nL(D_i) - \log(k + 1), \]  

where (72g) is due to the definition of \( L(\cdot) \) and (72h) is due to the fact that \( P_{e_i} \leq \epsilon_{i-1} \), for all \( 1 \leq i \leq k + 1 \).

This yields the lower bound in (28).

C. Proof of Theorem 2

We will need the following lemma in our proof for Theorem 2

**Lemma 4:** For any given \( n \) and \( f \), \( L^{(G)}(n, t, f) \) is convex in \( t \). Consequently, \( L^{(G)}(t, f) \) is also convex in \( t \), for any \( f \).

**Proof:** For any \( t_1, t_2 \), and some \( 0 \leq \lambda \leq 1 \), let \( t_{\lambda} = \lambda t_1 + (1 - \lambda) t_2 \). We will show that \( L(n, t_{\lambda}, f) \leq \lambda L(n, t_1, f) + (1 - \lambda) L(n, t_2, f) \). Let \( P_1 \) and \( P_2 \) be optimal mechanisms for \( L^{(G)}(n, t_1, f) \) and \( L^{(G)}(n, t_2, f) \) respectively, and \( P_{\lambda} \triangleq \lambda P_1 + (1 - \lambda) P_2 \). Note that \( P_{\lambda} \) is feasible for \( L^{(G)}(n, t_{\lambda}, f) \) because

\[
\mathbb{E}_{P_{\lambda}} \left[ f \left( d(X^n, \hat{X}^n) \right) \right] = \lambda \mathbb{E}_{P_1} \left[ f \left( d(X^n, \hat{X}^n) \right) \right] + (1 - \lambda) \mathbb{E}_{P_2} \left[ f \left( d(X^n, \hat{X}^n) \right) \right] \leq \lambda t_1 + (1 - \lambda) t_2 = t_{\lambda}.
\]

Moreover, since \( I(Y^n; \hat{X}^n) \) is convex in \( P_{\hat{X}^n|X^n, Y^n} \), the leakage achieved by \( P_{\lambda} \) is at most equal to \( \lambda L^{(G)}(n, t_1, f) + (1 - \lambda) L^{(G)}(n, t_2, f) \) which implies \( L^{(G)}(n, t_{\lambda}, f) \leq \lambda L^{(G)}(n, t_1, f) + (1 - \lambda) L^{(G)}(n, t_2, f) \). Finally we note that the asymptotic leakage \( L^{(G)}(t, f) \) is also convex in \( t \) because it is the limit of convex functions in \( t \).

We now present an achievable scheme and a converse for Theorem 2

**Achievability:** We know that \( L^{(G)}(t, f) \leq L^{(M)}(t, f) \leq L(f^{-1}(t)) \), where the latter inequality is due to Theorem 1. Since by Lemma 4, \( L^{(G)}(t, f) \) is a convex function in \( t \), the definition of lower convex envelope gives \( L^{(G)}(t, f) \leq (L \circ f^{-1})^{**}(t) \). This in turn gives \( L^{(G)}(t, f) \leq (L \circ f^{-1})^{**}(t) \) due to Remark 3.

**Converse:** We first focus on the class of piecewise step functions \( f \), and then show that the result holds for any function \( f \), using piecewise step approximations of \( f \).

**Piecewise Step functions \( f \):** Let us consider the class of functions \( f \) that are of the form

\[ f(D) = \sum_{i=1}^{k} a_i \mathbf{1}(D > D_i), \]  

where \( a_i \) are real numbers and \( D_i \) are real numbers such that \( D_1 < \ldots < D_k \).
where \( k \) is finite and each \( D_i \) is a distinct distortion level with \( f(D_i) < f(D_j) \) for \( i < j \). For this class of functions, (6) simplifies and can be lower bounded as

\[
L^{(G)}(n, t, f) = \min_{P_{X^n|X^n, Y^n}} \frac{1}{n} I(Y^n; \hat{X}^n)
\]

\[
= \min_{P_{X^n|X^n, Y^n}} \sum_{i=1}^{k} a_i 1(d(X^n, \hat{X}^n) > D_i) \leq t
\]

\[
= \min_{0 \leq \epsilon_1 \leq \cdots \leq \epsilon_t \leq 1: \sum_{i=1}^{k} a_i \epsilon_i \leq t} \min_{\gamma_i \geq 0, \gamma_i = 1 \iff D_i = D_{k+1}, \sum_{i=1}^{k+1} \gamma_i = 1} \sum_{i=1}^{k+1} \gamma_i L(D_i) - \frac{\log(k+1)}{n}
\]

\[
= \max_{\lambda \geq 0} \min_{\gamma_i \geq 0, \gamma_i = 1 \iff D_i = D_{k+1}, \sum_{i=1}^{k+1} \gamma_i = 1} \sum_{i=1}^{k+1} \gamma_i L(D_i) - \frac{\log(k+1)}{n}
\]

\[
= \max_{\lambda \geq 0} \min_{\gamma_i \geq 0} \sum_{i=1}^{k+1} \gamma_i L(D_i) - \frac{\log(k+1)}{n}
\]

\[
= \max_{\lambda \geq 0} \min_{f(D_i)} L(D_i) + \lambda f(D_i) - \frac{\log(k+1)}{n}
\]

\[
= \max_{\lambda \geq 0} \min_{f(D_i)} L(f^{-1}_u(t_i)) + \lambda t_i - \frac{\log(k+1)}{n}
\]

where

- (75c) follows from Lemma 1 and the fact that \( L(D_{\max}) = 0 \).
- (75d) is due to forming the Lagrangian given by incorporating only the last constraint in (75c), i.e. \( \sum_{i=1}^{k} a_i \epsilon_i \leq t \).
- (75e) is derived by letting \( \epsilon_{k+1} = 0, D_{k+1} = D_{\max}, \) and \( \gamma_i = \epsilon_{i-1} - \epsilon_i \), for \( i = 1, \ldots, k \).
- (75f) holds because a convex combination of non-negative real numbers is minimized by choosing a \( \gamma \) with \( \gamma_i = 1 \) for some \( i \) corresponding to the smallest \( L(D_i) + \lambda f(D_i) \), and \( \gamma_j = 0 \), for all other \( j \neq i \).
- (75g) is derived by defining \( t_i = f(D_i) \), i.e. \( D_i = f_u^{-1}(t_i) \).

Then, by taking the limit as \( n \to \infty \) we have

\[
L^{(G)}(t, f) = \max_{\lambda} \min_{i} L(f_u^{-1}(t_i)) + \lambda t_i - \lambda t.
\]
Note that the $i$th function, $L(f_u^{-1}(t_i)) + \lambda t_i$ is a minimizer for some $\lambda$, if for all $j \neq i$ we have

$$L(f_u^{-1}(t_i)) + \lambda t_i \leq L(f_u^{-1}(t_j)) + \lambda t_j,$$

or equivalently

$$\frac{L(f_u^{-1}(t_i)) - L(f_u^{-1}(t_j))}{t_i - t_j} \leq -\lambda, \quad \text{for } j < i,$$

(77a)

$$\frac{L(f_u^{-1}(t_i)) - L(f_u^{-1}(t_j))}{t_i - t_j} \geq -\lambda, \quad \text{for } j > i.$$  

(77b)

Note that (77a) and (77a) imply the slope of the line connecting points $\{(t_i, L(f_u^{-1}(t_i))), (t_j, L(f_u^{-1}(t_j)))\}$ is not larger than $-\lambda$, for $j < i$, and not smaller than $-\lambda$, for $j > i$. This holds if and only if $L(f_u^{-1}(t_i)) = (L \circ f_u^{-1})^\star\star(t_i)$. Since $L \circ f_u^{-1})^\star\star(t_i) = (L \circ f^{-1})^\star\star(t_i)$ due to Remark 3 the only relevant $i$ in the minimization in (78) are those for which $L(f_u^{-1}(t_i)) = (L \circ f^{-1})^\star\star(t_i)$. Hence, (78) can be rewritten as

$$L^{(G)}(t, f) \geq \max_{\lambda} \min_{i: L(f_u^{-1}(t_i)) = (L \circ f^{-1})^\star\star(t_i)} L(f_u^{-1}(t_i)) + \lambda t_i - \lambda t.$$  

(79)

For a chosen $\lambda$ and $i$, $L(f_u^{-1}(t_i)) + \lambda t_i - \lambda t$ is the evaluation of a linear function at $t$, which is tangential to $(L \circ f^{-1})^\star\star(\cdot)$ at $(t_i, (L \circ f^{-1})^\star\star(t_i))$, with slope $-\lambda$. This value is always smaller than or equal to $(L \circ f^{-1})^\star\star(t)$, and because $(L \circ f^{-1})^\star\star(\cdot)$ is a convex piecewise linear function, it suffices to optimize over only those values of $\lambda$ that are equal to the slope of the linear segment of $(L \circ f^{-1})^\star\star(\cdot)$ that contains $t$. Thus, for an optimal $\lambda$ we have $\min_i (L \circ f_u^{-1})(t_i) + \lambda t_i - \lambda t = (L \circ f^{-1})^\star\star(t)$, resulting in $L^{(G)}(t, f) \geq (L \circ f^{-1})^\star\star(t)$.

**General functions $f$**: Finally, we now show that $L^{(G)}(t, f) \geq (L \circ f^{-1})^\star\star(t)$ for the case of general non-decreasing left continuous functions $f$. For any $\delta > 0$, there exists a lower approximation $f_\delta$ of $f$ over $[D_{\min}, D_{\max}]$ that has the form of (74) with a finite number of step functions, i.e. $f_\delta(x) = \sum_{i=1}^k a_i 1_{D_i}(x)$, with $a_i = f(D_i) - f(D_{i-1})$ for $1 \leq i \leq k$ and $a_{\max} \triangleq \max_i a_i \leq \delta$. Then, we have $f_\delta(D) \leq f(D) \leq f_\delta(D) + \delta$, and thus

$$L^{(G)}(t, f) \geq L^{(G)}(t, f_\delta) \geq (L \circ f_\delta^{-1})^\star\star(t) = (L \circ f_\delta^{-1})^\star\star(t),$$

(80a)

(80b)

(80c)

where (80a) holds because we have $L^{(G)}(n, t, f_\delta) \leq L^{(G)}(n, t, f)$ for any $n$, (80b) is based on the result we had earlier on piecewise step functions specifically, and (80c) is due to Remark 3. Then, taking the limit as $\delta \to 0$ and the fact that $\lim_{\delta \to 0} f_\delta(D) = f(D)$ gives $L^{(G)}(t, f) \geq (L \circ f^{-1})^\star\star(t)$.

**D. Proof of Theorem 2**

We now proceed to proving the result in (22) for all non-increasing right-continuous functions $g : [D_{\min}, D_{\max}] \to (0, 1]$. Recall that we proved this for simple functions through Lemma 1. For any bounded, non-increasing, and right-continuous function $g$, there exists two sequences of simple functions $\{g_i\}_{i=1}^\infty$ and $\{g_\delta\}_{\delta=1}^\infty$ that are bounded away from zero, converge to $g$ uniformly from above and below, respectively, and each of functions $g_i$ and $g_\delta$ takes
i distinct values. Since \( g_i(D) \leq g(D) \leq \bar{g}_i(D) \) for all \( i \geq 1, D \in [D_{\min}, D_{\max}] \), and the asymptotic optimal leakage for simple constraint functions is the integral in (26), for each \( i \geq 1 \) we have
\[
\int_{D_{\min}}^{D_{\max}} L(D) d(\bar{g}_i(D)) \leq L(G) \leq \int_{D_{\min}}^{D_{\max}} L(D) d(g_i(D)).
\] (81)
Since \( L(\cdot) \) and \( g(\cdot) \) are bounded, the integral \( \int_{D_{\min}}^{D_{\max}} L(D) d(g_i(D)) \) exists. Therefore, in order to prove
\[
L(G) = \int_{D_{\min}}^{D_{\max}} L(D) d(g_i(D)),
\] (82)
it suffices to show that
\[
\lim_{i \to \infty} \int_{D_{\min}}^{D_{\max}} L(D) d(g_i(D)) = \int_{D_{\min}}^{D_{\max}} L(D) d(g_i(D)),
\] (83)
and similarly for the integral with respect to \( d(\bar{g}_i(D)) \). In order to do so, we use the uniform convergence of \( g_i \) to \( g \), and integration by parts. Since \( L(\cdot) \) is a convex, and therefore, continuous function, the Lebesgue–Stieltjes integral \( \int_{D_{\min}}^{D_{\max}} L(D) d(g_i(D)) \) reduces to a Riemann–Stieltjes integral, and admits integration by parts [24]. Thus, we can bound the difference of the two integrals as

\[
\left| \int_{D_{\min}}^{D_{\max}} L(D) d(g_i(D)) - \int_{D_{\min}}^{D_{\max}} L(D) d(g(D)) \right| \leq \left| L(D) \left( \frac{g_i(D)}{D_{\max}} - g(D) \right) \right|_{D_{\min}}^{D_{\max}} + \left| \int_{D_{\min}}^{D_{\max}} \left( g_i(D) - g(D) \right) d(L(D)) \right|,
\] (84a)

\[
= \left| L(D) \left( \frac{g_i(D)}{D_{\max}} - g(D) \right) \right|_{D_{\min}}^{D_{\max}} + \left| \int_{D_{\min}}^{D_{\max}} \left( g_i(D) - g(D) \right) d(L(D)) \right| \leq \left| L(D) \left( \frac{g_i(D)}{D_{\max}} - g(D) \right) \right|_{D_{\min}}^{D_{\max}} + \left| \int_{D_{\min}}^{D_{\max}} \left( g_i(D) - g(D) \right) d(L(D)) \right|,
\] (84b)

which goes to zero as \( i \to \infty \), due to uniform convergence of \( g_i \) to \( g \). One can also verify the same argument for \( d(\bar{g}_i(D)) \). Hence, both of the integrals in (81) converge to the same value \( \int_{D_{\min}}^{D_{\max}} L(D) d(g_i(D)) \), and therefore (82) holds.

E. Proof of Lemma 2

Due to the symmetry of the source distribution, and convexity of mutual information in conditional distribution, there exists an optimal mechanism with

\[
P(\hat{X} = 1|X = 0, Y = 1) = P(\hat{X} = 0|X = 1, Y = 0) = \beta_1,
\] (85)

\[
P(\hat{X} = 1|X = 0, Y = 0) = P(\hat{X} = 0|X = 1, Y = 1) = \beta_2.
\] (86)
Therefore, it suffices to optimize over all feasible values of $\beta_1$ and $\beta_2$. Rewriting the joint distribution $P_{Y,\hat{X}}$ in terms of $\beta_1$, $\beta_2$, and $q$ gives

\[
P(Y = 0, \hat{X} = 1) = P(Y = 1, \hat{X} = 0) = 0.5 \left[(1 - q)\beta_2 + q(1 - \beta_1)\right],
\]

\[
P(Y = 0, \hat{X} = 0) = P(Y = 1, \hat{X} = 1) = 0.5 \left[(1 - q)(1 - \beta_2) + q\beta_1\right].
\]

Therefore, we have

\[
L(D) = \min_{0 \leq \beta_1, \beta_2 \leq 1: (1-q)\beta_2 + q\beta_1 \leq D} H(\hat{X}) - H(\hat{X}|Y)
\]

\[
= \min_{0 \leq \beta_1, \beta_2 \leq 1: (1-q)\beta_2 + q\beta_1 \leq D} 1 - H_b((1 - q)\beta_2 + q(1 - \beta_1))
\]

\[
= \min_{q-D \leq \gamma \leq q+D, 0 \leq \gamma \leq 1} 1 - H_b(\gamma)
\]

\[
= \begin{cases} 
1 - H_b(q + D), & D < 0.5 - q, \\
0, & D \geq 0.5 - q.
\end{cases}
\]

where

- (89b) is due to (87) and (88).
- (89c) holds because $q \leq 0.5$ and the minimum and maximum values of $(1 - q)\beta_2 + q(1 - \beta_1)$ subject to $(1-q)\beta_2 + q\beta_1 \leq D$ are $\min\{q + D, 1\}$ and $\max\{q - D, 0\}$, respectively.

If $D < q$, then the extreme values occur at the corner points of the feasible region with $(\beta_1 = 0, \beta_2 = \frac{D}{1-q})$, and $(\beta_1 = \frac{D}{q}, \beta_2 = 0)$. Otherwise, if $q \leq D \leq 1 - q$, then the minimum and maximum values will be 0 and $q + D$, respectively. Finally, for $D > 1 - q$ the extreme values will be 0 and 1. The first scenario is depicted in Fig. 14.

- (89c) is due to the fact that the binary entropy function $H_b(\cdot)$ is concave and maximized at 0.5.

![Fig. 14. The feasible set and extreme values for $(1 - q)\beta_2 + q(1 - \beta_1)$ subject to $(1-q)\beta_2 + q\beta_1 \leq D$, if $D < q$.](image-url)
VI. CONCLUSION

We have formulated the tradeoff between privacy and utility as a minimization of mutual information between private and released data subject to two different forms of distortion constraints: the average distortion cost constraint and the complementary CDF bound on distortion. The former allows for taking non-separable distortion measures into account, while the latter enables the data publisher to provide refined guarantees on utility.

For the average distortion cost constraints, we have characterized the asymptotically optimal leakage for both stationary memoryless and general mechanisms as a function of the single letter leakage function \( L \) and the distortion cost function \( f \). In particular, we have shown that a memoryless mechanism achieves the asymptotically optimal leakage if and only if the information leakage-cost function \( L(f^{-1}()) \) coincides with its lower convex envelope; otherwise, a mixture of exactly two memoryless mechanisms is sufficient.

For the complementary CDF bound on distortion, we have derived the asymptotically optimal leakage. We have shown that under general mechanisms the optimal leakage is equal to the integral of the single letter leakage function with respect to the Lebesgue measure associated with the complementary CDF bound, while for stationary and memoryless mechanisms, it is equal to the single letter leakage function evaluated at the largest value of distortion for which the CDF bound function is equal to one.

For both types of utility constraints, the challenge remains to characterize the second order performance of the leakage as a function of the data size \( n \). More generally, the proof techniques developed here for arbitrary cost functions and complementary CDF bounds on distortion are applicable to a broad class of information theoretic problems such as lossy source coding with fidelity constraints and channel coding with input cost constraints.

REFERENCES

[1] K. Kalantari, L. Sankar, and O. Kosut, “On information-theoretic privacy with general distortion cost functions,” in 2017 IEEE International Symposium on Information Theory (ISIT), June 2017, pp. 2865–2869.
[2] R. W. Yeung, Information Theory and Network Coding, 1st ed. Springer Publishing Company, Incorporated, 2008.
[3] V. Kostina, Y. Polyanskiy, and S. Verdú, “Variable-length compression allowing errors,” IEEE Transactions on Information Theory, vol. 61, no. 8, pp. 4316–4330, Aug 2015.
[4] V. Kostina and S. Verdú, “Fixed-length lossy compression in the finite blocklength regime,” IEEE Trans. Inform. Theory, vol. 58, no. 6, pp. 3309–3338, Jun. 2012.
[5] I. Kontoyiannis and S. Verdú, “Optimal lossless data compression: Non-asymptotics and asymptotics,” Information Theory, IEEE Transactions on, vol. 60, no. 2, pp. 777–795, Feb 2014.
[6] Y. Shkel and S. Verdú, “A coding theorem for f-separable distortion measures,” in 2016 Information Theory and Applications Workshop (ITA), Jan 2016, pp. 1–7.
[7] Y. Aluğ, H. V. Poor, and S. Verdú, “On fixed-length channel coding with feedback in the moderate deviations regime,” in 2015 IEEE International Symposium on Information Theory (ISIT), June 2015, pp. 1816–1820.
[8] ———, “Variable-length channel codes with probabilistic delay guarantees,” in 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), Sept 2015, pp. 642–649.
[9] ———, “On channel dispersion per unit cost,” in 2016 IEEE International Symposium on Information Theory (ISIT), July 2016, pp. 2429–2433.
[10] K. Kalantari, O. Kosut, and L. Sankar, “On the fine asymptotics of information theoretic privacy,” in 2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2016.
[11] V. Y. F. Tan, “Asymptotic estimates in information theory with non-vanishing error probabilities,” Foundations and Trends in Communications and Information Theory, vol. 11, no. 1-2, pp. 1–184, 2014.
[12] H. Yamamoto, “A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers (corresp.),” *IEEE Transactions on Information Theory*, vol. 29, no. 6, pp. 918–923, Nov 1983.

[13] L. Sankar, S. R. Rajagopalan, and H. V. Poor, “Utility-privacy tradeoffs in databases: An information-theoretic approach,” *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 6, pp. 838–852, 2013.

[14] F. du Pin Calmon and N. Fawaz, “Privacy against statistical inference,” in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2012, pp. 1401–1408.

[15] F. du Pin Calmon, A. Makhdoumi, and M. Médard, “Fundamental limits of perfect privacy,” in *IEEE International Symposium on Information Theory, ISIT 2015, Hong Kong, China, June 14-19, 2015*, 2015, pp. 1796–1800. [Online]. Available: http://dx.doi.org/10.1109/ISIT.2015.7282765

[16] S. Asoodeh, M. Diaz, F. Alajaji, and T. Linder, “Information extraction under privacy constraints,” *CoRR*, vol. abs/1511.02381, 2015.

[17] K. Kalantari, L. Sankar, and A. D. Sarwate, “Optimal differential privacy mechanisms under hamming distortion for structured source classes,” in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 2069–2073.

[18] S. L. Warner, “Randomized response: A survey technique for eliminating evasive answer bias,” *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[19] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, “What can we learn privately?” *SIAM Journal on Computing*, vol. 40, no. 3, pp. 793–826, 2008.

[20] P. Kairouz, S. Oh, and P. Viswanath, “Extremal mechanisms for local differential privacy,” *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 492–542, Jan. 2016. [Online]. Available: http://dl.acm.org/citation.cfm?id=2946645.2946662

[21] J. Duchi, M. Jordan, and M. Wainwright, “Local privacy and statistical minimax rates,” in *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, Oct 2013, pp. 429–438.

[22] W. Hoeffding, “Probability inequalities for sums of bounded random variables,” *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 33–40, 1963. [Online]. Available: http://www.jstor.org/stable/2282952

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[24] E. Hille and R. S. Phillips, *Functional analysis and semi-groups*. American Mathematical Society, 1996, vol. 31.