Differential Privacy as a Mutual Information Constraint

Paul Cuff
Princeton University

Lanqing Yu
Princeton University

ABSTRACT

Differential privacy is a precise mathematical constraint meant to ensure privacy of individual pieces of information in a database even while queries are being answered about the aggregate. Intuitively, one must come to terms with what differential privacy does and does not guarantee. For example, the definition prevents a strong adversary who knows all but one entry in the database from further inferring about the last one. This strong adversary assumption can be overlooked, resulting in misinterpretation of the privacy guarantee of differential privacy.

Herein we give an equivalent definition of privacy using mutual information that makes plain some of the subtleties of differential privacy. The mutual-information differential privacy is in fact sandwiched between ϵ-differential privacy and (ϵ, δ)-differential privacy in terms of its strength. In contrast to previous works using unconditional mutual information, differential privacy is fundamentally related to conditional mutual information, accompanied by a maximization over the database distribution. The conceptual advantage of using mutual information, aside from yielding a simpler and more intuitive definition of differential privacy, is that its properties are well understood. Several properties of differential privacy are easily verified for the mutual information alternative, such as composition theorems.

Keywords

Differential privacy, information theory.

1. INTRODUCTION

Differential privacy is a concept proposed in [12] for database privacy. It allows queries to be answered about aggregate quantities of data while protecting the privacy of individual entries in the database. In the absence of a precise mathematical framework such as differential privacy, practitioners have been tempted to use various rules-of-thumb to protect privacy (e.g. "don't answer a query that averages fewer than k entries together"—see the query restriction approach in [1]). Instead, differential privacy directly addresses the statistical distinguishability of the database and has led to algorithms for answering general queries with just the right amount of randomness used in order to preserve privacy.[1]

Differential privacy requires that two adjacent databases, which differ in only one entry, are statistically indistinguishable, as measured by a probabilistic metric defined in Section 2. This guarantee is particularly effective for making individuals feel comfortable contributing personal information to a dataset. For instance, if a person decides to participate in a survey, his answers only constitute one response out of the entire collection, and the responses of other people remain unchanged. Differential privacy is meant to assure the one participant that his answers are concealed.

This privacy metric has gained a lot of traction in recent years. The main contribution of this work is to cast differential privacy as a mutual information constraint. There have been many attempts in the literature to connect differential privacy to mutual information. Here we give not only a connection but an equivalence.

To briefly summarize the main result, consider a database $X^n = (X_1, \ldots, X_n)$ that returns a query response Y according to a random mechanism $P_{Y|X^n}$. Let X^{-i} denote the set of database entries excluding X_i.

Definition 1 ((ϵ, δ)-Differential Privacy [13]). *A randomized mechanism $P_{Y|X^n}$ satisfies* (ϵ, δ)-differential privacy *if for all neighboring database instances x^n and $\tilde{x}^n$*

$$P_{Y|X^n=x^n} \overset{(\epsilon,\delta)}{\approx} P_{Y|X^n=\tilde{x}^n}, \tag{1}$$

where the approximation in (1) *is defined later in Definition 4, and neighboring database instances are defined in Definition 3 as any pair of database vectors that differ in only one entry (i.e. Hamming distance one).*[2]

Definition 2 (Mutual-Information Diff. Priv.). *A randomized mechanism $P_{Y|X^n}$ satisfies* ϵ-mutual-information differential privacy *if*

$$\sup_{i, P_{X^n}} I(X_i; Y|X^{-i}) \leq \epsilon \text{ nats}. \tag{2}$$

Note that nats *are the information units that result from using the natural logarithm instead of the logarithm base two, which would give bits.*

[1] Differential privacy does not assume the adversary has any computational limitation.

[2] Another similar definition for "neighbor" exists in the literature, involving the removal of one entry of the database.

CCS'16, October 24 - 28, 2016, Vienna, Austria

ISBN 978-1-4503-4139-4/16/10…$15.00

DOI: http://dx.doi.org/10.1145/2976749.2978308

The main claim of this paper, which appears in Section 3, is an equivalence between mutual-information differential privacy (MI-DP) and the standard definition of (ϵ, δ)-differential privacy ((ϵ, δ)-DP). The original definition of differential privacy [12], defined formally in Section 2, parameterized privacy with a single positive number ϵ. For various reasons it has since been relaxed [13] to have two parameters ϵ and δ playing multiplicative and additive roles in the likelihood constraint. We refer to the original DP as ϵ-DP and the relaxed form as (ϵ, δ)-DP. In this notation, ϵ-DP is simply $(\epsilon, 0)$-DP.

The claim herein is that MI-DP is sandwiched between these two definitions in the following sense: It is weaker than ϵ-DP but stronger than (ϵ, δ)-DP. That is, a mechanism that satisfies ϵ-DP also satisfies ϵ-MI-DP.[3] Similarly, if ϵ-MI-DP holds then (ϵ', δ)-DP also must hold, where ϵ' and δ vanish as ϵ goes to zero. In fact, the connection between MI-DP and (ϵ, δ)-DP is an equivalence if either the domain or range of the query mechanism is a finite set.

The advantage of this alternative but equally strong definition of differential privacy is that mutual information is a well-understood quantity. It provides a clear picture of what differential privacy does and does not guarantee. Furthermore, several properties of differential privacy are immediate to prove in this form.

While the mathematics of differential privacy, in its standard form, are straightforward, an intuitive understanding can be elusive. The definition of (ϵ, δ)-DP involves a notion of neighboring database instances. Upon examination one realizes that this has the affect of assuming that the adversary has already learned about all but one entry in the database and is only trying to gather additional information about the remaining entry. We refer to this as the strong adversary assumption, which is implicit in the definition of differential privacy. Notice that MI-DP needs no definition of neighborhood. The strong adversary assumption is made explicit in the conditioning within the conditional mutual information term.

The strong adversary assumption is both a feature and a vulnerability of the definition of differential privacy. It is a feature when recruiting individual participants for a survey. The individual can decide whether or not to participate but cannot do anything about the information contributed by others (which may inform on them indirectly). Differential privacy assures them that even with access to everyone else's responses, the survey reports will not further reveal anything about their individual response. However, DP also has its shortcomings as a privacy guarantee. Among all adversaries with different prior knowledge of the database, the strong adversary may not be the one which benefits the most from the query output. Indeed, it is shown in [18] that a weaker adversary can compromise privacy severely if the entries in the database are correlated, which is quite typical in certain applications such as social networks.

As an equivalent privacy metric, MI-DP benefits and suffers in the same way. Fortunately, the definition of MI-DP puts this potential weakness in plain sight. It shows explicitly that information leakage is only being restricted conditioned on the remainder of the database being known. Clearly, this does not bound the unconditional mutual information when correlations are present.

[3]The other direction is not true in general, as there exist mechanisms which satisfy ϵ-MI-DP but not ϵ'-DP for any ϵ'.

Somewhat paradoxically, mutual information can serve simultaneously as both a measure of privacy, as in MI-DP, and as a quantification of utility—for example, the mutual information between the entire database and the query response, $I(X^n; Y)$. The close connection between mutual information and estimation and detection further captures the privacy-utility trade-off.

Several works [20, 9, 4, 3, 10] relate mutual information to differential privacy by upper bounding mutual information given a differential privacy achieving mechanism. One common point of these works is that they all use unconditional mutual information rather than conditional. Often, the conclusion is a bound on the unconditional mutual information between the whole database and the private output. The use of conditional mutual information in Definition 2 captures the prior knowledge of the database possessed by a potential adversary (i.e. the strong adversary assumption implicit in DP). This is crucial in developing an equivalence with the standard DP definition.

Another crucial ingredient that some of the literature fails to properly incorporate (e.g. [25]) when applying mutual information to differential privacy is that differential privacy is a property of the query mechanism and assumes no specific prior distribution on the database. Mutual information, on the other hand, is not well defined without a joint distribution, which must include a distribution for the database. The remedy is to maximize the mutual information over all possible distributions on the database, as seen in the definition of MI-DP in Definition 2.[4] Had we defined MI-DP with respect to any particular database distribution (e.g. with independent and identically distributed entries), we would have significantly reduced its strength as a privacy metric. The formula in Definition 2 in fact looks like a channel capacity formula one would encounter in expressing the fundamental limit of communication through a noisy channel. This maximization removes any distributional assumption and makes MI-DP a property of the mechanism itself.

2. PRELIMINARIES

2.1 Notation

The set $\{1, 2, \cdots, m\}$ is denoted as $[m]$. An index set $\mathcal{I}$ is a subset of $[n]$ whose elements are enumerated as $(i_1, \cdots, i_{|\mathcal{I}|})$, where $|\cdot|$ denotes the cardinality of a set.

We use X^n as shorthand notation for the sequence of random variables $(X_1, \cdots, X_n)$. The symbol X^{-i} denotes the sequence of $n-1$ random variables $(X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n)$, in other words, all of X^n except X_i. The lower case symbol x^{-i} is an instance of X^{-i}. For any index set $\mathcal{I}$, we use $X_{\mathcal{I}}$ to denote the sequence of random variables $(X_{i_1}, \cdots, X_{i_{|\mathcal{I}|}})$ specified by $\mathcal{I}$. Similarly the lower case $x_{\mathcal{I}} = (x_{i_1}, \cdots, x_{i_{|\mathcal{I}|}})$ is an instance of $X_{\mathcal{I}}$.

A database X^n consists of n entries, where the i-th entry takes values from $\mathcal{X}_i$.

DEFINITION 3 (NEIGHBOR). *Two database instances x^n and $\tilde{x}^n$ are* neighbors *if they differ in only one entry. In other words,*

$$d_H(x^n, \tilde{x}^n) = 1, \tag{3}$$

where $d_H(\cdot, \cdot)$ is Hamming distance.

[4]This approach, using worst-case database distribution, appears in various works throughout the literature, e.g. [11].

The output of a privacy mechanism is a random variable represented as Y and takes values from $\mathcal{Y}$.

2.2 Statistical Indistinguishability

Two probability distributions can be considered statistically indistinguishable if they are close under an appropriate metric. The criterion for indistinguishability used in the standard definition of differential privacy is the following.

DEFINITION 4 ((ϵ,δ)-CLOSENESS). *Two probability distributions P and Q over the same measurable space $(\Omega,\mathcal{F})$ are (ϵ,δ)-close, denoted as*

$$P \overset{(\epsilon,\delta)}{\approx} Q \tag{4}$$

if

$$P(A) \leq e^{\epsilon} Q(A) + \delta, \quad \forall A \in \mathcal{F}, \tag{5}$$
$$Q(A) \leq e^{\epsilon} P(A) + \delta, \quad \forall A \in \mathcal{F}. \tag{6}$$

Consider two special cases, $\delta = 0$ and $\epsilon = 0$. If $\delta = 0$, then P and Q are mutually absolutely continuous, denoted as $P \lll\gg Q$, and $(\epsilon, 0)$-closeness is a statement about the Radon-Nikodym derivative $\frac{dP}{dQ}$:

$$P \overset{(\epsilon,0)}{\approx} Q \iff \left| \ln \frac{dP}{dQ}(a) \right| \leq \epsilon \quad \forall a \in \Omega. \tag{7}$$

On the other hand, with $\epsilon = 0$, $(0,\delta)$-closeness is a statement about the total variation distance:

$$P \overset{(0,\delta)}{\approx} Q \iff \|P - Q\|_{TV} \leq \delta. \tag{8}$$

We can also relate (ϵ,δ)-closeness to Kullback-Leibler divergence, denoted as $D(\cdot\|\cdot)$. For example, by relaxing the right side of (7) to be an expected value rather than a statement about all $a \in \Omega$, we immediately get the following implication:

$$P \overset{(\epsilon,0)}{\approx} Q \implies \begin{array}{l} D(P\|Q) \leq \epsilon \text{ nats}, \\ D(Q\|P) \leq \epsilon \text{ nats}. \end{array} \tag{9}$$

Tighter expressions of the relationship to Kullback-Leibler divergence are given next, in Properties 1 and 2. We give a proof of Property 1 in Appendix A.

PROPERTY 1.

$$P \overset{(\epsilon,0)}{\approx} Q \implies \begin{array}{l} D(P\|Q) \leq \min\left\{\epsilon, \epsilon^2\right\} \text{ nats}, \\ D(Q\|P) \leq \min\left\{\epsilon, \epsilon^2\right\} \text{ nats}. \end{array} \tag{10}$$

In fact, the tightest possible statement of this form is

$$P \overset{(\epsilon,0)}{\approx} Q \implies \begin{array}{l} D(P\|Q) \leq \epsilon \dfrac{(e^{\epsilon}-1)\left(1-e^{-\epsilon}\right)}{(e^{\epsilon}-1)+(1-e^{-\epsilon})} \text{ nats}, \\ D(Q\|P) \leq \epsilon \dfrac{(e^{\epsilon}-1)\left(1-e^{-\epsilon}\right)}{(e^{\epsilon}-1)+(1-e^{-\epsilon})} \text{ nats}. \end{array} \tag{11}$$

Equality on the right side of (11) *can be achieved with binary distributions. For small ϵ, the right side of* (11) *is asymptotically $\frac{1}{2}\epsilon^2$ nats.*

PROPERTY 2. *By Pinsker's inequality,*

$$D(P\|Q) \leq \epsilon \text{ nats} \implies P \overset{\left(0,\sqrt{\epsilon/2}\right)}{\approx} Q. \tag{12}$$

Property 1 and Property 2 are strict in the sense that the reverse implications are not true in any form (i.e. closeness bounds on the right, no matter how small the parameters, do not even imply finiteness of the parameters on the left). Also, we already mentioned that Property 1 is tight. Property 2 is known to be tight up to a multiplicative constant.

The quantities arising in the above definitions and properties have concrete connections to inference. Total variation distance precisely captures the error probability in a binary hypothesis test. That is, one minus the total variation distance is the minimum sum of the two types of binary error probability. Kullback-Leibler divergence precisely captures the asymptotic hypothesis testing error upon observing many independent observations [8, Chapter 11]. The strongest of these metrics, $(\epsilon, 0)$-closeness, has an interpretation in the Bayesian setting as a bound on the Bayes factor. That is, the log-posterior-odds-ratio cannot change by more than ϵ due to the observation. Finally, (ϵ,δ)-closeness is shown in [17] to be precisely a piecewise linear constraint on the error region in a binary hypothesis test. By inspection of that relationship, the following property is apparent (proven in Appendix B).

PROPERTY 3. *For any non-negative $\epsilon' < \epsilon$, let $\delta' = 1 - \frac{\left(e^{\epsilon'}+1\right)(1-\delta)}{e^{\epsilon}+1}$.*

$$P \overset{(\epsilon,\delta)}{\approx} Q \implies P \overset{(\epsilon',\delta')}{\approx} Q. \tag{13}$$

Property 3 is the tightest possible trade-off between ϵ and δ with respect to (ϵ,δ)-closeness. Notice that $\delta' > \delta$. It is not possible for a larger δ to imply a smaller one, for any finite ϵ and ϵ'.

2.3 Differential Privacy

The definition of (ϵ,δ)-DP in Definition 1 has been now made precise with Definition 3 (neighbor) and Definition 4 ((ϵ,δ)-closeness).

We define ϵ-DP and (δ)-DP by setting either of the two parameters to zero.

DEFINITION 5 (ϵ-DIFFERENTIAL PRIVACY [12]). *A randomized mechanism $P_{Y|X^n}$ satisfies ϵ-DP if it satisfies $(\epsilon, 0)$-DP.*

DEFINITION 6 ((δ)-DIFFERENTIAL PRIVACY). *A randomized mechanism $P_{Y|X^n}$ satisfies (δ)-DP if it satisfies $(0,\delta)$-DP.*

Mutual-information differential privacy was defined in the introduction in Definition 2.

Finally, let us define one additional privacy metric based on Kullback-Leibler divergence, which we will call KL-DP.

DEFINITION 7 (KL DIFFERENTIAL PRIVACY). *A randomized mechanism $P_{Y|X^n}$ satisfies ϵ-KL-DP if for all neighboring database instances x^n and $\tilde{x}^n$*

$$D\left(P_{Y|X^n=x^n} \middle\| P_{Y|X^n=\tilde{x}^n}\right) \leq \epsilon \text{ nats}. \tag{14}$$

2.4 Ordering of Privacy Metrics

This work is about showing equivalence of privacy metrics. In order to do so, we must define an ordering.

Definition 8 (Stronger Privacy Metric). *As a placeholder, take α-DP and β-DP to represent two generic privacy guarantees with positive parameters α and β. We say that α-DP is stronger than β-DP, denoted as*

$$\alpha\text{-DP} \succeq \beta\text{-DP}, \tag{15}$$

if for all $\beta' > 0$ there exists an $\alpha' > 0$ such that

$$\alpha'\text{-DP} \implies \beta'\text{-DP}. \tag{16}$$

If the parameters are vectors, then $\beta' > 0$ and $\alpha' > 0$ should be interpreted as inequalities on each coordinate.

Example 1. *It is clear that ϵ-DP $\succeq$ (ϵ, δ)-DP and (δ)-DP $\succeq$ (ϵ, δ)-DP, since ϵ-DP implies (ϵ, δ)-DP for any non-negative δ, and likewise for (δ)-DP, by definition.*

Also, (ϵ, δ)-DP $= (\delta)$-DP by Property 3. Notice that even if we set $\epsilon' = 0$, the quantity δ', as defined in the property, goes to zero as ϵ and δ go to zero.

3. MAIN RESULT

3.1 Equivalence

The emphasis of this work is the equivalence of mutual-information differential privacy with classical differential privacy.

Theorem 1 (Main Result).

$$\epsilon\text{-DP} \succeq \text{MI-DP} \succeq (\epsilon, \delta)\text{-DP}. \tag{17}$$

Furthermore, if the cardinality of the database entries or the query response is bounded, then

$$\text{MI-DP} = (\epsilon, \delta)\text{-DP}, \tag{18}$$

where the relationship (ϵ, δ)-DP $\succeq$ MI-DP is dependent on the cardinality bound

$$\min\left\{|\mathcal{Y}|, \max_i |\mathcal{X}_i|\right\}. \tag{19}$$

Precise bounds for the privacy parameters are given in the three lemmas in Section 3.3.

3.2 Related Work

Using information theoretic measures to quantify the privacy guarantee of differential privacy is not a new idea. An upper bound of mutual information is shown in [20] in a two-party differential privacy setting. Later this upper bound is used in [9] to get $I(X^n; Y) \leq 3\epsilon n$. In [3] and [4], min-entropy is considered rather than the usual Shannon entropy, and upper bounds are proven. In fact, [4, Corollary 1] implies an ordering relationship similar to the first inequality of (17) but for min-entropy based information leakage with only a single database entry. In [25], a "mutual information privacy" metric is defined and studied.

These works have in common that they all consider the use of unconditional mutual information. This doesn't capture the structure in the definition of differential privacy and the bounds are limited to the mutual information between the whole database and the sanitized query output, with no focus on individual entries. Needless to say, an equivalence is not established.

Some of the information theory literature bares resemblance to this work. In [5], similar proof steps to this work are used to show an equivalence between semantic security and a mutual information constraint. As in this work, there is a maximization over distributions of inputs to the randomized mechanism; however, conditional mutual information is not a part of that result, while it is a necessary ingredient here. Also, the notion of (ϵ, δ)-closeness goes by the name of E_γ distance in some of the information theory literature, such as [22] and [19]. Specifically, $P \overset{(\epsilon,\delta)}{\approx} Q$ is equivalent to the pair of statements $E_{e^\epsilon}(P\|Q) \leq \delta$ and $E_{e^\epsilon}(Q\|P) \leq \delta$.

3.3 Proof of Theorem 1

We prove (17) of Theorem 1 by proving a stronger chain of inequalities:

$$\epsilon\text{-DP} \overset{\text{(A)}}{\succeq} \text{KL-DP} \overset{\text{(B)}}{\succeq} \text{MI-DP} \overset{\text{(C)}}{\succeq} (\delta)\text{-DP} \overset{\text{(D)}}{=} (\epsilon, \delta)\text{-DP}. \tag{20}$$

It is worth noting both (A) and (B) are in fact strict orderings ($\succ$)—the reverse implications do not hold, even if cardinality bounds are assumed.

We now state the components of the proof in separate lemmas. Orderings (A) and (B) are the subject of Lemma 1, and ordering (C) is handled by Lemma 2. Equality (D) comes from Property 3, as discussed in Example 1.

Lemma 1 (Orderings (A) and (B)).

$$\epsilon\text{-DP} \implies \min\left\{\epsilon, \epsilon^2\right\}\text{-KL-DP}, \tag{21}$$

$$\epsilon\text{-KL-DP} \implies \epsilon\text{-MI-DP}. \tag{22}$$

Therefore,

$$\epsilon\text{-DP} \implies \min\left\{\epsilon, \epsilon^2\right\}\text{-MI-DP}. \tag{23}$$

Proof. The first statement, (21), is established by Property 1. Both ϵ-DP and KL-DP are defined the same way in terms of neighboring database instances.

The second statement, (22), is best understood through the geometric interpretation of capacity as the radius of the information ball [8, Theorem 13.1.1]. The radius cannot be more than the maximum of pairwise distances. However, we will not directly use that machinery here. Instead, consider the following direct proof.

Start by assuming that the randomized mechanism $P_{Y|X^n}$ satisfies ϵ-KL-DP. Let $i \in \{1, \ldots, n\}$ and P_{X^n} be arbitrary. For notational clarity, let $\bar{X}^n \sim P_{X^n}$, and begin with a representation of conditional mutual information for a general distribution in terms of Kullback-Leibler divergence:

$$I(X_i; Y|X^{-i}) = \mathbb{E}\left[D\left(P_{Y|X^n=\bar{X}^n}\middle\|P_{Y|X^{-i}=\bar{X}^{-i}}\right)\right] \tag{24}$$

Now we bound $D\left(P_{Y|X^n=x^n}\middle\|P_{Y|X^{-i}=x^{-i}}\right)$ for each instance x^n. Fix x^n arbitrarily, and let $\tilde{X} \sim P_{X_i|X^{-i}=x^{-i}}$. Consider,

$$P_{Y|X^{-i}=x^{-i}} = \mathbb{E}\left[P_{Y|X_i=\tilde{X}, X^{-i}=x^{-i}}\right]. \tag{25}$$

Therefore, by Jensen's inequality, and using the fact that $D(\cdot\|\cdot)$ is convex in the second argument, we conclude,

$$\begin{aligned} &D\left(P_{Y|X^n=x^n}\middle\|P_{Y|X^{-i}=x^{-i}}\right) \\ &= D\left(P_{Y|X^n=x^n}\middle\|\mathbb{E}\left[P_{Y|X_i=\tilde{X}, X^{-i}=x^{-i}}\right]\right) \\ &\leq \mathbb{E}\left[D\left(P_{Y|X^n=x^n}\middle\|P_{Y|X_i=\tilde{X}, X^{-i}=x^{-i}}\right)\right] \\ &\leq \epsilon \text{ nats}, \end{aligned} \tag{26}$$

where the last inequality is due to the fact that any two databases that agree on X^{-i} are neighbors. □

LEMMA 2 (ORDERING (C)).

$$\epsilon\text{-MI-DP} \implies \left(0, \sqrt{2\epsilon}\right)\text{-DP}. \tag{27}$$

In fact, the tightest possible statement of this form is

$$\epsilon\text{-MI-DP} \implies \left(0, \delta'\right)\text{-DP}, \tag{28}$$

with $\delta' = 1 - 2h^{-1}(\ln 2 - \epsilon)$*, where* h^{-1} *is the inverse of the increasing part of the binary entropy function in units of nats. This formula holds for* $\epsilon \in [0, \ln 2]$*. For* $\epsilon > \ln 2$*, the implication becomes* (1)*-DP, which is vacuous.*

The claim in (27) *is looser than that in* (28) *but asymptotically tight for small* ϵ*.*

PROOF. The essence of this claim is found in the binary case, with a binary database and a binary query response. We show this reduction first.

Start by assuming that the randomized mechanism $P_{Y|X^n}$ satisfies ϵ-MI-DP. Consider an arbitrary pair of neighboring database instances x^n and $\tilde{x}^n$, and let i be the location where they differ. Denote by $\Delta_{x^n,\tilde{x}^n}$ the subset of probability distributions over the space of databases $\mathcal{D}$ that only put positive mass on x^n and $\tilde{x}^n$. Therefore, all distributions in $\Delta_{x^n,\tilde{x}^n}$ are binary, and X^{-i} is deterministic with respect to any of them.

Also, let A be an arbitrary measurable subset of $\mathcal{Y}$. Consider the indicator function

$$B(y) = \begin{cases} 1, & y \in A, \\ 0, & y \notin A. \end{cases} \tag{29}$$

The random variable B is the binary function $B(Y)$.

$$\begin{aligned} \max_{P_{X^n} \in \Delta_{x^n,\tilde{x}^n}} I(X_i; B) &\overset{(a)}{\leq} \max_{P_{X^n} \in \Delta_{x^n,\tilde{x}^n}} I(X_i; Y) \\ &\overset{(b)}{=} \max_{P_{X^n} \in \Delta_{x^n,\tilde{x}^n}} I(X_i; Y|X^{-i}) \\ &\leq \sup_{P_{X^n}} I(X_i; Y|X^{-i}) \\ &\overset{(c)}{\leq} \epsilon \text{ nats}, \end{aligned} \tag{30}$$

where (a) is due to the data processing inequality, (b) comes from the fact that X^{-i} is deterministic for all distributions in $\Delta_{x^n,\tilde{x}^n}$, and (c) is by assumption of ϵ-MI-DP.

To summarize, we have arrived at a binary input and binary output randomized mechanism $P_{B|X_i}$, where $X_i \in \{x_i, \tilde{x}_i\}$, defined by

$$P_{B|X_i=x_i}(\{1\}) = P_{Y|X^n=x^n}(A), \tag{31}$$

$$P_{B|X_i=\tilde{x}_i}(\{1\}) = P_{Y|X^n=\tilde{x}^n}(A). \tag{32}$$

This mechanism is shown in (30) to satisfy ϵ-MI-DP. Also, since A, x^n, and $\tilde{x}^n$ were chosen arbitrarily, any (δ)-DP claim that can be made about $P_{B|X_i}$ must also hold for $P_{Y|X^n}$.

In Appendix C, we complete the proof by showing that Lemma 2 holds for all randomized mechanisms with a binary input and binary output, and that the characterization is tight.

A more complete characterization is also possible, of the form

$$\epsilon\text{-MI-DP} \implies \left(\epsilon', \delta'\right)\text{-DP}, \tag{33}$$

for a particular set of values (ϵ', δ') which, among other things, has the property that δ' must be greater than some positive threshold which depends on ϵ, and as δ' approaches this threshold, ϵ' must go to infinity. This characterization is also arrived at by first reducing to the binary case as we have done above. However, a description of the trade-off is too unwieldy for this discussion. □

We prove (18) of Theorem 1 with the following claim. The proof is given in Appendix D.

LEMMA 3 (REVERSE DIRECTION). *If* $|\mathcal{X}_i|$ *is finite for all* $i \in \{X_1, \ldots, X_n\}$*, or if* $|\mathcal{Y}|$ *is finite, then*

$$(\delta)\text{-DP} \implies \epsilon'\text{-MI-DP}, \tag{34}$$

where, for any $\delta \in [0, 1]$*,*

$$\epsilon' = 2h(\delta) + 2\delta \ln\left(\min\left\{|\mathcal{Y}|, \max_i |\mathcal{X}_i| + 1\right\}\right). \tag{35}$$

Slightly tighter bounds can be found in (105) and (126) of the proof. Although these bounds may have some looseness, the following example shows that they get roughly within a factor of two of the correct scaling for large cardinalities.

EXAMPLE 2 (ERASURE CHANNEL). *Consider a database with only one entry,* X_1*. Let* $\mathcal{X}_1 = [N]$ *and* $\mathcal{Y} = [N] \cup \{0\}$*. Define*

$$P_{Y|X_1=x_1} = \begin{cases} 1-\delta, & y = 0, \\ \delta, & y = x_1, \\ 0, & \text{otherwise.} \end{cases} \tag{36}$$

This randomized mechanism is usually referred to as an erasure channel, where the output $Y = 0$ *is considered an erasure. It is known that the capacity of this channel is*

$$C = \delta \log N, \tag{37}$$

where $N = |\mathcal{X}_1| = |\mathcal{Y}| - 1$*. This implies that there exists a distribution of the database (in this case, the uniform distribution for* $X_1 \in \mathcal{X}_1$*) such that*

$$\begin{aligned} I(X_1; Y|X^{-1}) &= I(X_1; Y) \\ &= \delta \log |\mathcal{X}_1| \\ &= \delta \log(|\mathcal{Y}| - 1). \end{aligned} \tag{38}$$

4. PROPERTIES OF DIFF. PRIVACY

Now that we have MI-DP as an equivalent metric of privacy, we explore the insights that this brings and simple proofs of properties about privacy.

The following are three basic and well-known properties of mutual information:

PROPERTY 4. *If* U *is independent of* W*, then*

$$I(U; V|W) \geq I(U; V). \tag{39}$$

PROPERTY 5. *If* U*,* V*, and* W *form a Markov chain* $U - V - W$*, meaning that* U *and* W *are conditionally independent given* V*, then*

$$I(U; V|W) \leq I(U; V). \tag{40}$$

PROPERTY 6 (CHAIN RULE).

$$I(U; V, W) = I(U; V) + I(U; W|V). \tag{41}$$

We will use these three properties (sometimes conditioned on other random variables) to make claims about MI-DP.

4.1 The Strong Adversary Assumption

We refer to a strong adversary as one who knows the entire database except for any one entry X_i. Differential privacy is implicitly designed as a protection against further information leakage to this adversary. The definition of MI-DP, now shown to be equivalent, makes this attribute explicit by conditioning on the remainder of the database and bounding $I(X_i;Y|X^{-i})$. But how much information does the sanitized output Y leak to an adversary with no prior knowledge?

In [18], this is referred to as *evidence of participation*. In the mutual information context, this may be measured by the unconditional mutual information $I(X_i;Y)$. It is pointed out in [18] that if the entries of the database are independent, the evidence of participation can be protected properly by differential privacy. This claim is straightforward using MI-DP in light of Property 4.

COROLLARY 1 (INDEPENDENT DATA). *If $\{X_i\}_{i=1}^n$ are mutually independent and $P_{Y|X^n}$ satisfies ϵ-MI-DP, then*

$$\sup_{i,P_{X_i}P_{X^{-i}}} I(X_i;Y) \leq \sup_{i,P_{X_i}P_{X^{-i}}} I(X_i;Y|X^{-i}) \leq \epsilon \text{ nats.} \tag{42}$$

On the other hand, it is often the case that entries of a database are correlated. Differential privacy does not provide a strong guarantee about the evidence of participation in general. Consider the following familiar example:

EXAMPLE 3 (CORRELATED DATABASE). *Consider a database with n binary entries. A data curator decides to release the mean of all entries and chooses the Laplace mechanism. Noise with distribution $\text{Lap}(\frac{1}{n\epsilon})$ is added to the sample mean to ensure ϵ-DP (also ϵ-MI-DP by Lemma 1).*

Now suppose all database entries are in fact equal to each other (maximally correlated). Let $X \sim \text{Bern}(0.5)$ and $X_i = X$ for all $i \in \{1,\ldots,n\}$. For large enough n, the noise added is negligible, and the binary value of the sample mean can be estimated with high accuracy, revealing each individual entry. In terms of mutual information, $I(X_i;Y) \approx 1$ bit for each i even while $I(X_i;Y|X^{-i}) = 0$ because $H(X_i|X^{-i}) = 0$.

4.2 Composition

Among the most important properties of differential privacy is composability. This states that a collection of queries, each satisfying differential privacy, collectively satisfies differential privacy with a parameter scaled proportional to the number of queries.

A great deal of effort has been made in deriving tight composition theorems for differential privacy. A straightforward composition theorem can be found in [13]. More intricate trade-offs can be found in [14] and [17], with the latter establishing a tight characterization.

The following claims for MI-DP mirror those found in [21] for $(\epsilon,0)$-DP and are in fact tight.

COROLLARY 2 (CONDITIONALLY INDEPENDENT QUERIES). *If several query responses $\{Y_1,\ldots,Y_k\}$ are produced conditionally independently given the database, and each mechanism $P_{Y_j|X^n}$ satisfies ϵ_j-MI-DP individually, then as a collection $P_{Y^k|X^n}$ satisfies $\left(\sum_j \epsilon_j\right)$-MI-DP.*

PROOF. For any i and P_{X^n}, the chain rule of mutual information (Property 6) gives (a), and Property 5 gives (b):

$$\begin{aligned} I(X_i;Y^k|X^{-1}) &\overset{(a)}{=} \sum_{j=1}^k I(X_i;Y_j|X^{-i},Y^{j-1}) \\ &\overset{(b)}{\leq} \sum_{j=1}^k I(X_i;Y_j|X^{-i}) \\ &\leq \sum_{j=1}^k \epsilon_j \text{ nats.} \end{aligned} \tag{43}$$

□

Corollary 2 states that the effect of releasing multiple conditionally independent query responses has no more than an additive effect on the parameter of privacy. It is worth noting two important points. First, query responses that are not conditionally independent (i.e. the noise from one query response is somehow reused in the next) have no such guarantee, as the following example illustrates.

EXAMPLE 4 (CORRELATED QUERY RESPONSES). *Consider a database where each entry has a finite alphabet $|\mathcal{X}_i| \leq \infty$. Consider two outputs of a query mechanism, $Y_1 = X_1 \oplus U$ and $Y_2 = U$, where U is a uniformly distributed random variable on the set $\{1,\ldots,|\mathcal{X}_1|\}$, independent of the database instance, and $\oplus$ is addition modulo $|\mathcal{X}_1|$. In other words, the first output Y_1 is X_1 encrypted by a one-time pad, and the second output Y_2 is the key to the one-time pad. Clearly, the combination of Y_1 and Y_2 reveals X_1 and violates differential privacy.*

On the other hand, Example 4 does not imply that correlated query responses should not be considered. Quite to the contrary, query responses that are carefully constructed to be correlated with each other have the potential to achieve significantly better privacy after multiple queries, as demonstrated in [15] and [6].

In general, the same composition claim of Corollary 2 holds even if the query responses are correlated as long as each response in sequence is specifically designed to satisfy differential privacy even with respect to the previous responses. The following corollary states this claim, and the proof follows directly from the proof of Corollary 2 simply by skipping (43).

COROLLARY 3 (SEQUENTIAL QUERIES). *If several query responses $\{Y_1,\ldots,Y_k\}$ are produced in sequence, and each mechanism $P_{Y_j|X^n,Y^{j-1}}$ satisfies ϵ_j-MI-DP individually, then as a collection $P_{Y^k|X^n}$ satisfies $\left(\sum_j \epsilon_j\right)$-MI-DP.*

The next claim is about query responses that each depend on different subsets of the database.

COROLLARY 4 (PARTIAL QUERIES). *If several query responses $\{Y_1,\ldots,Y_k\}$ are produced conditionally independently of each other from disjoint subsets of the database entries, denoted as $X_{\mathcal{I}_1},\ldots,X_{\mathcal{I}_k}$, with each mechanism $P_{Y_j|X_{\mathcal{I}_j}}$ satisfying ϵ-MI-DP individually, then as a collection $P_{Y^k|X^n}$ also satisfies ϵ-MI-DP.*

PROOF. Let $f(i)$ be the index j such that $i \in \mathcal{I}_j$. For any i and P_{X^n}, the chain rule of mutual information (Property 6)

gives:

$$\begin{aligned} I(X_i; Y^k|X^{-1}) &= I(X_i; Y_{f(i)}|X^{-i}) + I(X_i; Y^{-f(i)}|X^{-i}, Y_{f(i)}) \\ &= I(X_i; Y_{f(i)}|X^{-i}) \\ &\leq \epsilon \text{ nats.} \end{aligned} \tag{44}$$

□

5. A DISCREPANCY

While most properties of ϵ-DP or (ϵ, δ)-DP are also properties of MI-DP, it turns out that one basic property does not carry over.

Differential privacy is defined with respect to neighboring database instances. What privacy can be guaranteed if some bounded number of entries are changed in the database? Similar to the composition properties, the closeness of the output distribution scales proportionally with the number of database changes. The following properties are obtained by repeated application (5) and (6) from Definition 4.

PROPERTY 7 (EPSILON). *Suppose x^n and $\tilde{x}^n$ are instances of the database that differ in at most k entries, and that the randomized mechanism $P_{Y|X^n}$ is ϵ-DP. Then*

$$P_{Y|X^n=x^n} \overset{(k\epsilon,0)}{\approx} P_{Y|X^n=\tilde{x}^n}. \tag{45}$$

PROPERTY 8 (DELTA). *Suppose x^n and $\tilde{x}^n$ are instances of the database that differ in at most k entries, and that the randomized mechanism $P_{Y|X^n}$ is (δ)-DP. Then*

$$P_{Y|X^n=x^n} \overset{(0,k\delta)}{\approx} P_{Y|X^n=\tilde{x}^n}. \tag{46}$$

PROPERTY 9 (GENERAL). *Suppose x^n and $\tilde{x}^n$ are instances of the database that differ in at most k entries, and that the randomized mechanism $P_{Y|X^n}$ is (ϵ, δ)-DP. Then*

$$P_{Y|X^n=x^n} \overset{\left(k\epsilon, \frac{e^{k\epsilon}-1}{e^{\epsilon}-1}\delta\right)}{\approx} P_{Y|X^n=\tilde{x}^n}. \tag{47}$$

On the other hand, MI-DP does not have an analogous property. Even if a mechanism satisfies ϵ-MI-DP, there may not be a bound on $I(X_{\mathcal{I}}; Y|X_{\mathcal{I}^c})$, where $\mathcal{I}$ represents a subset of $|\mathcal{I}| = k$ indices. Consider the following example.

EXAMPLE 5. *Consider a database with two entries, X_1 and X_2, which are real valued. The randomized mechanism $P_{Y|X_1,X_2}$ produces an output Y which can be a real number or one of two special values e_1 or e_2. The behavior of the mechanism is best described in two cases:*

If $X_1 = X_2$:

$$Y = \begin{cases} X_1, & \text{with probability } \epsilon, \\ e_1, & \text{with probability } 1-\epsilon. \end{cases} \tag{48}$$

If $X_1 \neq X_2$:

$$Y = \begin{cases} e_2, & \text{with probability } \epsilon, \\ e_1, & \text{with probability } 1-\epsilon. \end{cases} \tag{49}$$

This mechanism satisfies $(\epsilon \ln 2)$-MI-DP. Notice that for any value of $X_2 = x_2$, we have a binary erasure channel from X_1 to Y, with binary input determined by whether $X_1 = x_2$ or not. The symbol e_1 serves as the erasure. The symbol e_2 represents the unerased indicator that $X_1 \neq x_2$. This binary erasure channel with erasure probability $1-\epsilon$ has mutual information bounded above by $\epsilon \ln 2$ nats (the capacity of the erasure channel).

On the other hand, the mutual information $I(X_1, X_2; Y)$ is unbounded if there are no constraints on $\mathcal{X}_1$ and $\mathcal{X}_2$. Indeed, if we let X_1 be a continuous random variable, and we set $X_2 = X_1$, then

$$I(X_1, X_2; Y) = \infty. \tag{50}$$

More generally, if the domains $\mathcal{X}_1$ and $\mathcal{X}_2$ are equal, then the capacity of the erasure channel gives the achievable mutual information (where e_2 represents an additional input symbol selected by any choice of $X_1 \neq X_2$):

$$\begin{aligned} \max_{P_{X_1,X_2}} I(X_1, X_2; Y) &= \epsilon \log(|\mathcal{X}_1| + 1) \\ &= \epsilon \log(|\mathcal{Y}| - 1). \end{aligned} \tag{51}$$

In fact, Example 5 might be best interpreted as a fortunate advantage of MI-DP. With any query mechanism, there is a trade-off between privacy and the informational utility to be gained from the output. If we apply Property 7 with $k = n$, the conclusion is that

$$P_{Y|X^n=x^n} \overset{(n\epsilon,0)}{\approx} P_{Y|X^n=\tilde{x}^n} \tag{52}$$

for any two databases x^n and $\tilde{x}^n$. By revisiting the proof of Lemma 1, we obtain

$$I(X^n; Y) \leq \min\left\{n\epsilon, (n\epsilon)^2\right\} \text{ nats.} \tag{53}$$

One way to view this is as a crude bound on the utility of the query output. The bound is detrimental if n is not large. On the other hand, MI-DP does not imply such a constraint.

If, however, we take into account a cardinality bound on the database entries or the query output, then there is indeed an upper bound on the information leaked from a group of database entries. This is obtained by using Property 8 in combination with Lemma 2, followed by repeating the proof of Lemma 3 for a group rather than an individual entry.

COROLLARY 5. *Suppose the randomized mechanism $P_{Y|X^n}$ satisfies ϵ-MI-DP. Then for any subset of indices $\mathcal{I}$, with $|\mathcal{I}| = k$, and with $\mathcal{I}^c = [n] \setminus \mathcal{I}$,*

$$\sup_{P_{X^n}} I(X_{\mathcal{I}}; Y|X_{\mathcal{I}^c}) \leq 2h\left(k\sqrt{2\epsilon}\right) + 2k\sqrt{2\epsilon}\log M, \tag{54}$$

where $M = \min\left\{|\mathcal{Y}|, (\max_i |\mathcal{X}_i|)^k + 1\right\}$.

6. VARIATIONS OF DIFF. PRIVACY

Many variations of differential privacy have been proposed in the literature to provide different assurances. Here we demonstrate how mutual-information differential-privacy can be adapted to correspond to these various definitions.

6.1 Personalized Differential Privacy

Personalized differential privacy [16] addresses the situation where participants of the database may have different concerns about the level of privacy. This is handled by assigning a different ϵ_i for each database entry X_i. That is, for any database instances x^n and $\tilde{x}^n$ which differ in only the ith place,

$$P_{Y|X^n=x^n} \overset{(\epsilon_i,0)}{\approx} P_{Y|X^n=\tilde{x}^n}. \tag{55}$$

The modification to MI-DP would be to require that for each i,

$$\sup_{P_{X^n}} I(X_i; Y|X^{-i}) \leq \epsilon_i \text{ nats.} \tag{56}$$

6.2 Free-Lunch Privacy

Free-lunch privacy was both defined and refuted in [18] as a stronger privacy definition which puts no restriction on which database instances must be indistinguishable. A mechanism $P_{Y|X^n}$ is ϵ-free-lunch private if every pair of database instances x^n and $\tilde{x}^n$ satisfies

$$P_{Y|X^n=x^n} \overset{(\epsilon,0)}{\approx} P_{Y|X^n=\tilde{x}^n}. \tag{57}$$

The MI-DP equivalent of this would be

$$\sup_{P_{X^n}} I(X^n; Y) \leq \epsilon \text{ nats.} \tag{58}$$

We can easily see the strength of this definition by applying the chain rule of mutual information (Property 6) to (58). The result is that for any pair of disjoint index sets $\mathcal{I}$ and $\mathcal{J}$,

$$\sup_{P_{X^n}} I(X_\mathcal{I}; Y|X_\mathcal{J}) \leq \epsilon \text{ nats.} \tag{59}$$

On the other hand, (58) and (59) illustrate the poor utility provided by the ϵ-free-lunch privacy mechanism, as the information contained in the output is always upper bounded by ϵ regardless of distribution and prior knowledge.

6.3 Bayesian Differential Privacy

Bayesian differential privacy [26] deals with the possible privacy degradation of differential privacy if the entries in the database are correlated. As was discussed in Section 4.1, a weak adversary who has less background knowledge of the database may stand to gain much more information than the adversary who knows all but one entry. Bayesian differential privacy is meant to protect simultaneously against all adversaries, but in order to do so it assumes a prior distribution on the database.

Given a prior distribution P_{X^n}, a mechanism $P_{Y|X^n}$ is ϵ-Bayesian differentially private if, for any index i and subset of indices $\mathcal{I}$,

$$P_{Y|X_i=x_i, X_\mathcal{I}=x_\mathcal{I}} \overset{(\epsilon,0)}{\approx} P_{Y|X_i=\tilde{x}_i, X_\mathcal{I}=x_\mathcal{I}}. \tag{60}$$

Notice that the conditional distributions are not necessarily conditioned on the entire database.

The MI-DP equivalent is, for any index i and subset of indices $\mathcal{I}$,

$$I(X_i; Y|X_\mathcal{I}) \leq \epsilon \text{ nats,} \tag{61}$$

which is in fact implied by (60). Furthermore, this notion of privacy can be strengthened by maximizing over database distributions, making it a stronger notion of privacy than differential privacy.

In spite of the additional strength of this privacy metric (especially when removing the Bayesian prior assumption by maximizing over the database distribution), this is not nearly as pessimistic as free-lunch privacy. As a comparison, the chain rule of mutual information (Property 6) in this case implies that for any two disjoint index sets $\mathcal{I}$ and $\mathcal{J}$,

$$I(X_\mathcal{I}; Y|X_\mathcal{J}) \leq |\mathcal{I}|\epsilon \text{ nats.} \tag{62}$$

Consequently,

$$I(X^n; Y) \leq n\epsilon \text{ nats.} \tag{63}$$

6.4 Adversarial Privacy

Adversarial privacy [23] does three things differently from differential privacy. First, it assumes a prior distribution P_{X^n} on the database (like Bayesian differential privacy). Second, it does not restrict attention to neighboring database instances (like free-lunch privacy). Third, it asymmetrically requires

$$\ln \frac{dP_{X^n|Y=y}}{dP_{X^n}} \leq \epsilon \quad \forall y \in \mathcal{Y}. \tag{64}$$

The idea is that the adversary can not increase certainty about a particular database value by much, even while other database values may be eliminated.

Since mutual information is the expected value of the quantity on the left of (64), it is clear that adversarial privacy implies

$$I(X^n; Y) \leq \epsilon \text{ nats.} \tag{65}$$

Thus, adversarial privacy has similarity to free-lunch privacy, though in the Bayesian setting. The subtleties of the asymmetric constraint are not captured in this MI-DP variant.

7. RÉNYI ENTROPY GENERALIZATION

The notion of α-mutual-information is the generalization of mutual information using Rényi information measures. There are many proposed ways to accomplish such a generalization. Here we adopt Sibson's proposal (see [24]):

$$I_\alpha^\mathsf{s}(X; Y) = \min_{Q_Y} D_\alpha(P_{Y|X} \| Q_Y | P_X), \tag{66}$$

where D_α is the conditional Rényi divergence of order α and the minimization is over all distributions Q_Y on $\mathcal{Y}$. Shannon's mutual information corresponds to $\alpha = 1$.

A simple upper bound holds for α-mutual information for all $\alpha \geq 0$, given in the following lemma.

LEMMA 4 (α-MUTUAL-INFORMATION UPPER BOUND). *If a randomized mechanism $P_{Y|X^n}$ satisfies ϵ-DP, then for all $\alpha \geq 0$, i, and instances of the remainder of the database x^{-i},*

$$\sup_{P_{X_i}} I_\alpha^\mathsf{s}(X_i; Y|X^{-i} = x^{-i}) \leq \epsilon \text{ nats.} \tag{67}$$

PROOF. Let $\alpha \geq 0$, i, and P_{X^n} be arbitrary. In order to abbreviate notation, denote the event $\{X^{-i} = x^{-i}\}$ as U. Pick an arbitrary $x_i \in \mathcal{X}_i$,

$$\begin{aligned} I_\alpha^\mathsf{s}(X_i; Y|U) &= \min_{Q_Y} D_\alpha(P_{Y|X_i,U} \| Q_Y | P_{X_i|U}) \\ &\leq D_\alpha(P_{Y|X_i,U} \| P_{Y|X_i=x_i,U} | P_{X_i|U}) \\ &= D_\alpha(P_{Y|X_i,U} P_{X_i|U} \| P_{Y|X_i=x_i,U} P_{X_i|U}) \\ &= \frac{1}{\alpha-1} \log \mathbb{E}\left[\frac{dP_{Y|X_i,U} P_{X_i|U}}{dP_{Y|X_i=x_i,U} P_{X_i|U}} (Y^*, X^*) \right]^{\alpha-1} \\ &= \frac{1}{\alpha-1} \log \mathbb{E}\left[\frac{dP_{Y|X_i,U}}{dP_{Y|X_i=x_i,U}} (Y^*, X^*) \right]^{\alpha-1}, \end{aligned} \tag{68}$$

where $(Y^*, X^*) \sim P_{Y|X_i,U} P_{X_i|U}$.

For $\alpha \neq 1$, the ϵ-DP constraint implies that $\frac{dP_{Y|X_iU}}{dP_{Y|X_i=x_i,U}} \leq e^\epsilon$ for all values of X_i, thus

$$\begin{aligned} I_\alpha^\mathsf{s}(X_i; Y|U) &\leq \frac{1}{\alpha-1} \log \mathbb{E}[e^\epsilon]^{\alpha-1} \\ &= \epsilon \text{ nats.} \end{aligned} \tag{69}$$

For the case $\alpha = 1$, the α-mutual-information reduces to Shannon's mutual information, and we have $I^{\mathsf{s}}_\alpha(X_i;Y|U) = I(X_i;Y|U) \leq \epsilon$ from Lemma 1. $\square$

Combining Property 7 with the proof of Lemma 4 gives the following corollary:

COROLLARY 6. *If the mechanism $P_{Y|X^n}$ satisfies ϵ-DP, then*

$$\sup_{P_{X^n}} I^{\mathsf{s}}_\alpha(X^n;Y) \leq n\epsilon \text{ nats.} \tag{70}$$

Furthermore, for $\alpha > 0$, when maximizing over database distributions P_{X^n}, all three notions of α-mutual-information discussed in [24] are equivalent. Thus,

$$\sup_{P_{X^n}} I^{\mathsf{s}}_\alpha(X^n;Y) = \sup_{P_{X^n}} I^{\mathsf{a}}_\alpha(X^n;Y) = \sup_{P_{X^n}} I^{\mathsf{c}}_\alpha(X^n;Y) \leq n\epsilon \text{ nats.} \tag{71}$$

In [3] and [4], the information leakage is defined as

$$I_\infty(X^n;Y) = H_\infty(X^n) - H_\infty(X^n|Y) \tag{72}$$

where

$$H_\infty(X^n|Y) = -\log \mathbb{E}\left[\max_{x^n} P_{X^n|Y}(x^n|Y)\right]. \tag{73}$$

This definition matches Arimoto's proposal $I^{\mathsf{a}}_\infty(X^n;Y)$, so it is a special case of (71).

8. ACKNOWLEDGEMENTS

This work was supported by the Air Force Office of Scientific Research (grant FA9550-15-1-0180) and the National Science Foundation (grant CCF-1350595).

9. REFERENCES

[1] N. R. Adam and J. C. Worthmann. Security-control methods for statistical databases: A comparative study. *ACM Comput. Surv.*, 21(4):515–556, Dec. 1989.

[2] R. Alicki and M. Fannes. Continuity of quantum conditional information. *Journal of Physics A: Mathematical and General*, 37(5):L55, 2004.

[3] M. S. Alvim, M. E. Andrés, K. Chatzikokolakis, P. Degano, and C. Palamidessi. Differential privacy: on the trade-off between utility and information leakage. In *Formal Aspects of Security and Trust*, pages 39–54. Springer Berlin Heidelberg, 2012.

[4] G. Barthe and B. Köpf. Information-theoretic bounds for differentially private mechanisms. In *24th Computer Security Foundations Symposium (CSF)*, pages 191–204. IEEE, 2011.

[5] M. Bellare, S. Tessaro, and A. Vardy. Semantic security for the wiretap channel. In *Advances in Cryptology–CRYPTO*, pages 294–311. Springer, 2012.

[6] A. Blum, K. Ligett, and A. Roth. A learning theory approach to noninteractive database privacy. *J. ACM*, 60(2):12:1–12:25, May 2013.

[7] H. Boche, R. F. Schaefer, and H. V. Poor. On the continuity of the secrecy capacity of compound and arbitrarily varying wiretap channels. *IEEE Transactions on Information Forensics and Security*, 10(12):2531–2546, Dec 2015.

[8] T. Cover and J. A. Thomas. *Elements of information theory*. Hoboken, NJ: Wiley-Interscience, 2 edition, 2006.

[9] A. De. Lower bounds in differential privacy. In *Theory of Cryptography*, pages 321–338. Springer, 2012.

[10] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and statistical minimax rates. In *54th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 429–438. IEEE, 2013.

[11] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Privacy aware learning. *J. ACM*, 61(6):38:1–38:57, Dec. 2014.

[12] C. Dwork. Differential privacy. *Automata, Languages and Programming*, pages 1–12, 2006.

[13] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology-EUROCRYPT*, pages 486–503. Springer, 2006.

[14] C. Dwork, G. N. Rothblum, and S. Vadhan. Boosting and differential privacy. In *51st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 51–60. IEEE, Oct 2010.

[15] M. Hardt and G. N. Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *51st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 61–70. IEEE, Oct 2010.

[16] Z. Jorgensen, T. Yu, and G. Cormode. Conservative or liberal? personalized differential privacy. In *31st International Conference on Data Engineering*, pages 1023–1034. IEEE, April 2015.

[17] P. Kairouz, S. Oh, and P. Viswanath. The composition theorem for differential privacy. In *32nd International Conference on Machine Learning*, 2015.

[18] D. Kifer and A. Machanavajjhala. No free lunch in data privacy. In *SIGMOD Int'l. Conference on Management of data*, pages 193–204. ACM, 2011.

[19] J. Liu, P. Cuff, and S. Verdú. Resolvability in e_γ with applications to lossy compression and wiretap channels. In *Int'l. Symp. on Information Theory (ISIT)*, pages 755–759. IEEE, 2015.

[20] A. McGregor, I. Mironov, T. Pitassi, O. Reingold, K. Talwar, and S. Vadhan. The limits of two-party differential privacy. In *51st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 81–90. IEEE, 2010.

[21] F. D. McSherry. Privacy integrated queries: an extensible platform for privacy-preserving data analysis. In *SIGMOD International Conference on Management of data*, pages 19–30. ACM, 2009.

[22] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. on Information Theory*, 56(5):2307–2359, 2010.

[23] V. Rastogi, M. Hay, G. Miklau, and D. Suciu. Relationship privacy: output perturbation for queries with joins. In *28th SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 107–116. ACM, 2009.

[24] S. Verdú. α-mutual information. In *Information Theory and Applications Workshop*, 2015.

[25] W. Wang, L. Ying, and J. Zhang. On the relation between identifiability, differential privacy, and mutual-information privacy. In *52nd Annual Allerton*

Conference on Communication, Control, and Computing (Allerton), pages 1086–1092, Sept 2014.

[26] B. Yang, I. Sato, and H. Nakagawa. Bayesian differential privacy on correlated data. In *SIGMOD International Conference on Management of Data*, pages 747–762. ACM, 2015.

[27] Z. Zhang. Estimating mutual information via kolmogorov distance. *IEEE Transactions on Information Theory*, 53(9):3280–3282, 2007.

APPENDIX

A. PROOF OF PROPERTY 1

Assume that

$$P \overset{(\epsilon,0)}{\approx} Q. \tag{74}$$

As stated in (7), this gives

$$\left|\ln \frac{dP}{dQ}(a)\right| \leq \epsilon \quad \forall a \in \Omega, \tag{75}$$

which is equivalent to

$$\frac{dP}{dQ}(a) \in \left[e^{-\epsilon}, e^{\epsilon}\right] \quad \forall a \in \Omega. \tag{76}$$

Consider that

$$\begin{aligned} D(P\|Q) &= \int dP(a) \ln \frac{dP}{dQ}(a) \\ &= \int dQ(a) \frac{dP}{dQ}(a) \ln \frac{dP}{dQ}(a) \\ &= \mathbb{E}\left[\frac{dP}{dQ}(X) \ln \frac{dP}{dQ}(X)\right], \end{aligned} \tag{77}$$

where $X \sim Q$.

Let us define the random variable $Z = \frac{dP}{dQ}(X)$. We know the following facts:

$$Z \in \left[e^{-\epsilon}, e^{\epsilon}\right] \quad \text{w.p. } 1, \tag{78}$$

$$\mathbb{E}[Z] = 1, \tag{79}$$

$$D(P\|Q) = \mathbb{E}[Z \ln Z]. \tag{80}$$

Since the function $f(x) = x \ln x$ for $x > 0$ is convex, we know that a distribution of Z that maximizes $D(P\|Q)$ under these constraints places all mass at the endpoints of the allowed support interval. Therefore, maximum $D(P\|Q)$ occurs with the following choice of distribution for Z:

$$Z = \begin{cases} e^{\epsilon}, & \text{w.p. } \frac{1-e^{-\epsilon}}{e^{\epsilon}-e^{-\epsilon}}, \\ e^{-\epsilon}, & \text{w.p. } \frac{e^{\epsilon}-1}{e^{\epsilon}-e^{-\epsilon}}. \end{cases} \tag{81}$$

A computation of $\mathbb{E}[Z \ln Z]$ gives the desired result.

This extreme is achieved by a symmetric pair of binary distributions, consistent with the distribution of Z derived above. Thus, coincidentally, for this choice of extreme distributions that maximize $D(P\|Q)$, it turns out that $D(P\|Q) = D(Q\|P)$.

The relaxation in Property 1 can be arrived at by making the following observation:

$$\begin{aligned} \epsilon \frac{(e^{\epsilon}-1)\left(1-e^{-\epsilon}\right)}{(e^{\epsilon}-1)+(1-e^{-\epsilon})} &\leq \epsilon\left(1-e^{-\epsilon}\right) \\ &\leq \min\left\{\epsilon, \epsilon^2\right\}. \end{aligned} \tag{82}$$

Other bounds in the literature (Lemma III.2 of [14] and Theorem 1 of [10]), while slightly loose, establish that $(\epsilon, 0)$-closeness implies an upper bound of roughly ϵ^2 nats of Kullback-Leibler divergence for small ϵ, which is only off by a factor of two. □

B. PROOF OF PROPERTY 3

Assume the P and Q are (ϵ, δ)-close. To show that they are (ϵ', δ')-close, we must show that for any $A \in \mathcal{F}$

$$P(A) \leq \delta' + e^{\epsilon'} Q(A), \tag{83}$$

$$Q(A) \leq \delta' + e^{\epsilon'} P(A). \tag{84}$$

By symmetry, we need only argue (83).

We will build the proof from two inequalities. The first is a direction application of (5):

$$P(A) \leq \delta + e^{\epsilon} Q(A). \tag{85}$$

The second is an application of (6) to the complement of A, denoted as A^c:

$$Q(A^c) \leq \delta + e^{\epsilon} P(A^c). \tag{86}$$

By substituting $P(A^c) = 1 - P(A)$ and $Q(A^c) = 1 - Q(A)$ and rearranging, this implies

$$P(A) \leq 1 - e^{-\epsilon}(1-\delta) + e^{-\epsilon} Q(A). \tag{87}$$

Now we complete the proof with some simple manipulations and by substituting the value $\delta' = 1 - \frac{\left(e^{\epsilon'}+1\right)(1-\delta)}{e^{\epsilon}+1}$ stated in Property 3. From (85) we can conclude

$$\begin{aligned} P(A) &\leq \delta + e^{\epsilon} Q(A) \\ &= \delta' + e^{\epsilon'} Q(A) + (\delta - \delta') + \left(e^{\epsilon} - e^{\epsilon'}\right) Q(A) \\ &= \delta' + e^{\epsilon'} Q(A) + \left(e^{\epsilon} - e^{\epsilon'}\right)\left(Q(A) - \frac{1-\delta}{e^{\epsilon}+1}\right). \end{aligned} \tag{88}$$

From (87) we have

$$\begin{aligned} P(A) &\leq 1 - e^{-\epsilon}(1-\delta) + e^{-\epsilon} Q(A) \\ &= \delta' + e^{\epsilon'} Q(A) + \left(1 - e^{-\epsilon}(1-\delta) - \delta'\right) + \left(e^{-\epsilon} - e^{\epsilon'}\right) Q(A) \\ &= \delta' + e^{\epsilon'} Q(A) + \left(e^{\epsilon'} - e^{-\epsilon}\right)\left(\frac{1-\delta}{e^{\epsilon}+1} - Q(A)\right). \end{aligned} \tag{89}$$

If $Q(A) \leq \frac{1-\delta}{e^{\epsilon}+1}$ then (88) establishes (85), since $\epsilon \geq \epsilon' \geq 0$. Otherwise, (89) establishes (85). □

C. PROOF OF LEMMA 2

According to the arguments immediately following Lemma 2, we only need to show that the claim holds for randomized mechanisms $P_{Y|X}$ that have binary input and binary output. That is, $|\mathcal{X}| = |\mathcal{Y}| = 2$.

Start by assuming that the randomized mechanism $P_{Y|X}$ satisfies ϵ-MI-DP. Since X is a database with only one entry, ϵ-MI-DP simply means

$$\max_{P_X} I(X;Y) \leq \epsilon \text{ nats}. \tag{90}$$

Notice that the left side is the expression for channel capacity from information theory, where $P_{Y|X}$ would be interpreted as a communication channel. With this interpretation, what we are trying to show is that a bound on the channel capacity for binary channels implies a total variation bound between the conditional output distributions. It has already been argued that binary channels contain the extreme cases, since total variation can be expressed as an inequality relating probabilities of a single arbitrary set (in general, the form of (5) and (6) gives this conclusion). The next step is to show, specifically for total variation, that binary *symmetric* channels are the extreme cases.

Because X and Y are binary, the channel $P_{Y|X}$ can be parametrized with two parameters:

$$a \triangleq \mathbb{P}[Y=1|X=0], \tag{91}$$

$$b \triangleq \mathbb{P}[Y=1|X=1]. \tag{92}$$

Now consider the complementary channel $P_{\tilde{Y}|\tilde{X}}$, where $\tilde{X} = X \oplus 1$ and $\tilde{Y} = Y \oplus 1$, with $\oplus$ representing addition modulo 2. This gives

$$\mathbb{P}\left[\tilde{Y}=1\middle|\tilde{X}=0\right] = 1-b, \tag{93}$$

$$\mathbb{P}\left[\tilde{Y}=1\middle|\tilde{X}=1\right] = 1-a. \tag{94}$$

Finally, define a new binary channel, denoted as $P_{\hat{Y}|\hat{X}}$, which is a convex combination of the original channel and the complementary channel. Then,

$$\mathbb{P}\left[\hat{Y}=1\middle|\hat{X}=0\right] = \frac{1}{2} + \frac{a-b}{2}, \tag{95}$$

$$\mathbb{P}\left[\hat{Y}=1\middle|\hat{X}=1\right] = \frac{1}{2} - \frac{a-b}{2}. \tag{96}$$

Notice that for all three channels, $P_{Y|X}$, $P_{\tilde{Y}|\tilde{X}}$, and $P_{\hat{Y}|\hat{X}}$, the total variation between the two conditional output distributions is the same:

$$\begin{aligned}\left\|P_{Y|X=0} - P_{Y|X=1}\right\|_{TV} &= \left\|P_{\tilde{Y}|\tilde{X}=0} - P_{\tilde{Y}|\tilde{X}=1}\right\|_{TV} \\ &= \left\|P_{\hat{Y}|\hat{X}=0} - P_{\hat{Y}|\hat{X}=1}\right\|_{TV} \\ &= |a-b|. \end{aligned} \tag{97}$$

On the other hand, channel capacity is a convex function of the channel parameters. By symmetry, $P_{Y|X}$ and $P_{\tilde{Y}|\tilde{X}}$ have the same capacity. Therefore, the convex combination $P_{\hat{Y}|\hat{X}}$, which is a binary *symmetric* channel, has a lower capacity. Thus, binary symmetric channels are the extreme points in the trade-off between capacity and total variation. For every binary channel, there is a binary symmetric channel with the same capacity but with greater or equal total variation distance between the conditional output distributions.

Finally, we arrive at Lemma 2 by applying the formula for channel capacity of a binary symmetric channel. If we denote by δ the total variation distance between the conditional output distributions, then the cross-over probability is $\frac{1}{2} - \frac{\delta}{2}$. The channel capacity is then

$$C = \ln 2 - h\left(\frac{1}{2} - \frac{\delta}{2}\right) \text{ nats}, \tag{98}$$

where $h(\cdot)$ is the binary entropy function in nats. Inverting this equation gives (28).

The relaxed bound in (27) is established by the fact that the second order Tailor expansion of $h(x)$ about $x = \frac{1}{2}$ is in fact an upper bound:

$$h(x) \le \ln 2 - 2\left(x - \frac{1}{2}\right)^2. \tag{99}$$

An alternative simple argument directly arrives at the looser bound in (27), without even reducing to the binary case. We again refer to the geometric interpretation of capacity as the radius of the information ball [8, Theorem 13.1.1]. By Pinsker's inequality (Property 2), each conditional output distribution is within total variation distance $\sqrt{\frac{\epsilon}{2}}$ of the center of the information ball. The triangle inequality gives (27). □

D. PROOF OF LEMMA 3

Assume that the randomized mechanism $P_{Y|X^n}$ is (δ)-DP, and let $i \in \{1,\dots,n\}$ and P_{X^n} be arbitrary.

Two proof arguments are needed, one based on the database entries $\{X_i\}$ having a finite set of possible values, and the other based on the same for the query response Y. In both cases, however, we first note that the conditional mutual information $I(X_i; Y|X^{-i})$ is an expected value over instances of X^{-i}. We provide bounds that uniformly hold for each instance of x^{-i}. To that end, fix x^{-i} arbitrarily, and let $(\tilde{X},\tilde{Y}) \sim P_{X_i,Y|X^{-i}=x^{-i}}$. This gives,

$$I(X_i; Y|X^{-i}=x^{-i}) = I(\tilde{X};\tilde{Y}). \tag{100}$$

Notice further that any two databases in the set $\{\tilde{x}^n : \tilde{x}^{-i} = x^{-i}\}$ are neighbors according to Definition 3. Therefore, by assumption,

$$P_{\tilde{Y}|\tilde{X}=\tilde{x}_1} \overset{(0,\delta)}{\approx} P_{\tilde{Y}|\tilde{X}=\tilde{x}_2} \tag{101}$$

for any two values $\tilde{x}_1$ and $\tilde{x}_2$.

We now aim to bound $I(\tilde{X};\tilde{Y})$. Consider first the case where $|\mathcal{Y}| < \infty$. By construction, $|\tilde{\mathcal{Y}}| = |\mathcal{Y}|$.

From (101) we can claim that for any value $\tilde{x}$

$$P_{\tilde{Y}|\tilde{X}=\tilde{x}} \overset{(0,\delta)}{\approx} P_{\tilde{Y}}. \tag{102}$$

This is justified by letting $X' \sim P_{\tilde{X}}$ and noting

$$\begin{aligned}\left\|P_{\tilde{Y}|\tilde{X}=\tilde{x}} - P_{\tilde{Y}}\right\|_{TV} &= \left\|P_{\tilde{Y}|\tilde{X}=\tilde{x}} - \mathbb{E}\left[P_{\tilde{Y}|\tilde{X}=X'}\right]\right\|_{TV} \\ &\le \mathbb{E}\left[\left\|P_{\tilde{Y}|\tilde{X}=\tilde{x}} - P_{\tilde{Y}|\tilde{X}=X'}\right\|_{TV}\right] \\ &\le \delta, \end{aligned} \tag{103}$$

where the first inequality is due to Jensen's inequality and the convexity of the total variation distance.

Next we decompose mutual information into entropy terms:

$$I(\tilde{X};\tilde{Y}) = H(\tilde{Y}) - H(\tilde{Y}|\tilde{X}). \tag{104}$$

Finally, a continuity property of entropy found in [27] (see (4) within), derived from optimal coupling and Fano's inequality, bounds the difference in entropy as a function of total variation distance and $|\tilde{\mathcal{Y}}|$. Combining this with (102) and (104) gives

$$I(\tilde{X};\tilde{Y}) \le \begin{cases} h(\delta) + \delta \ln\left(|\tilde{\mathcal{Y}}|-1\right), & \delta \le \frac{|\tilde{\mathcal{Y}}|-1}{|\tilde{\mathcal{Y}}|}, \\ \ln|\tilde{\mathcal{Y}}|, & \delta > \frac{|\tilde{\mathcal{Y}}|-1}{|\tilde{\mathcal{Y}}|} \end{cases} \tag{105}$$

$$\le h(\delta) + \delta \ln|\tilde{\mathcal{Y}}| \text{ nats}. \tag{106}$$

Next we consider the case where $\max_i |\mathcal{X}_i| < \infty$. By construction, $|\tilde{\mathcal{X}}| \le \max_i |\mathcal{X}_i|$.

For this case, we take (102) a bit further. In fact,

$$P_{\tilde{X},\tilde{Y}} \overset{(0,\delta)}{\approx} P_{\tilde{X}} P_{\tilde{Y}}. \tag{107}$$

This is justified by letting $X' \sim P_{\tilde{X}}$ and noting

$$\left\| P_{\tilde{X},\tilde{Y}} - P_{\tilde{X}} P_{\tilde{Y}} \right\|_{TV} = \mathbb{E}\left[\left\| P_{\tilde{Y}|\tilde{X}=X'} - P_{\tilde{Y}} \right\|_{TV} \right]. \tag{108}$$

This time we decompose mutual information in the reverse direction:

$$I(\tilde{X};\tilde{Y}) = H(\tilde{X}) - H(\tilde{X}|\tilde{Y}). \tag{109}$$

To complete the proof we need a continuity argument for condition entropy. The following lemma is inspired by ideas from [7] which in turn come from [2].

Lemma 5 (Continuity of conditional entropy). *If P and Q are two distributions on $\mathcal{U} \times \mathcal{V}$ with $|\mathcal{U}| < \infty$, then*

$$\begin{gathered} P \overset{(0,\delta)}{\approx} Q \\ \Downarrow \\ |H_P(U|V) - H_Q(U|V)| \le 2h\left(\frac{\delta}{\delta+1}\right) + 2\frac{\delta}{\delta+1}\log|\mathcal{U}|. \end{gathered} \tag{110}$$

Proof. Since the bound in the lemma is monotonic in δ, we assume without loss of generality that

$$\|P - Q\|_{TV} = \delta. \tag{111}$$

We first translate closeness in total variation distance to the existence of a common distribution that is close to both relative to the boundaries of the set of probability distributions. To be more precise, there exists a probability distribution p^* which is a convex combination of P and another probability distribution, with most of the convex weight on P, and the same relationship holds between p^* and Q. That is:

$$p^* = \frac{1}{1+\delta}P + \frac{\delta}{1+\delta}\hat{p} \tag{112}$$

$$= \frac{1}{1+\delta}Q + \frac{\delta}{1+\delta}\hat{q}. \tag{113}$$

Once we have established this existence, the exact construction of p^*, $\hat{p}$, and $\hat{q}$ will have no consequence on the conclusion.

Consider the Hahn decomposition of the signed measure $P - Q$ into positive and negative parts that are mutually singular, represented by the non-negative measures μ^+ and μ^-, as follows:

$$P - Q = \mu^+ - \mu^-, \tag{114}$$

$$\mu^+ \ge 0, \tag{115}$$

$$\mu^- \ge 0, \tag{116}$$

$$\mu^+ \perp \mu^-. \tag{117}$$

The total measure of each part, μ^+ and μ^-, is the total variation between P and Q, which is δ. Thus, to normalize μ^+ and μ^- to become probability measures, we must divide by δ.

Let

$$\hat{p} = \frac{1}{\delta}\mu^-, \tag{118}$$

$$\hat{q} = \frac{1}{\delta}\mu^+. \tag{119}$$

Then we have that p^* is the greater part of P and Q, normalized as

$$p^* = \frac{1}{1+\delta}\left(P + \mu^-\right) \tag{120}$$

$$= \frac{1}{1+\delta}\left(Q + \mu^+\right), \tag{121}$$

which satisfies both (112) and (113).

Next, to complete the proof, we show that

$$|H_P(U|V) - H_{p^*}(U|V)| \le h\left(\frac{\delta}{\delta+1}\right) + \frac{\delta}{\delta+1}\log|\mathcal{U}|, \tag{122}$$

$$|H_Q(U|V) - H_{p^*}(U|V)| \le h\left(\frac{\delta}{\delta+1}\right) + \frac{\delta}{\delta+1}\log|\mathcal{U}|. \tag{123}$$

By symmetry, an argument for only one of the inequalities is needed.

The following bound is aided by defining a binary random variable $B \sim \text{Bern}\left(\frac{\delta}{1+\delta}\right)$ from which we construct $p^*_{U,V|B=0} = P$ and $p^*_{U,V|B=1} = \hat{p}$. This has no effect on the marginal distribution of $p^*_{U,V}$. We have

$$\begin{aligned} H_{p^*}(U|V) &= H_{p^*}(U|V,B) + I_{p^*}(U;B|V) \\ &= \frac{1}{1+\delta}H_P(U|V) + \frac{\delta}{1+\delta}H_{\hat{p}}(U|V) + I_{p^*}(U;B|V). \end{aligned} \tag{124}$$

Subtracting $H_P(U|V)$ from both sides gives

$$\begin{aligned} &H_{p^*}(U|V) - H_P(U|V) \\ &= \frac{\delta}{1+\delta}\left(H_{\hat{p}}(U|V) - H_{p^*}(U|V)\right) + I_{p^*}(U;B|V). \end{aligned} \tag{125}$$

Finally, the argument is completed by bounding the three non-negative terms. The entropy terms are bounded by $\log|\mathcal{U}|$. For the conditional mutual information, $I_{p^*}(U;B|V) \le H_{p^*}(B) = h\left(\frac{\delta}{\delta+1}\right)$. □

The proof of Lemma 3 is completed by applying Lemma 5 with $P_{\tilde{X},\tilde{Y}}$ as P and $P_{\tilde{X}}P_{\tilde{Y}}$ as Q, due to (107). Combined with (109) this gives

$$I(\tilde{X};\tilde{Y}) \le 2h\left(\frac{\delta}{\delta+1}\right) + 2\frac{\delta}{\delta+1}\ln|\tilde{\mathcal{X}}| \tag{126}$$

$$\le 2h(\delta) + 2\delta\ln\left(|\tilde{\mathcal{X}}| + 1\right) \text{ nats.} \quad \square \tag{127}$$