MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

On Privacy-Utility Tradeoffs for Constrained Data Release Mechanisms

Basciftci, Yuksel Ozan; Wang, Ye; Ishwar, Prakash

TR2016-019 February 01, 2016

Abstract

Privacy-preserving data release mechanisms aim to simultaneously minimize information-leakage with respect to sensitive data and distortion with respect to useful data. Dependencies between sensitive and useful data results in a privacy-utility tradeoff that has strong connections to generalized rate-distortion problems. In this work, we study how the optimal privacy-utility tradeoff region is affected by constraints on the data that is directly available as input to the release mechanism. In particular, we consider the availability of only sensitive data, only useful data, and both (full data). We show that a general hierarchy holds, that is, the tradeoff region given only the sensitive data is no larger than the region given only the useful data, which in turn is clearly no larger than the region given both sensitive and useful data. In addition, we determine conditions under which the tradeoff region given only the useful data coincides with that given full data. This is based on the common information between the sensitive and useful data.

IEEE Information Theory and Applications Workshop (ITA) 2016

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

On Privacy-Utility Tradeoffs for Constrained Data Release Mechanisms

Yuksel Ozan Basciftci
The Ohio State University
Columbus, Ohio
Email: basciftci.1@osu.edu

Ye Wang
Mitsubishi Electric Research Laboratories
Cambridge, Massachusetts
Email: yewang@merl.com

Prakash Ishwar
Boston University
Boston, Massachusetts
Email: pi@bu.edu

***Abstract*—Privacy-preserving data release mechanisms aim to simultaneously minimize information-leakage with respect to sensitive data and distortion with respect to useful data. Dependencies between sensitive and useful data results in a privacy-utility tradeoff that has strong connections to generalized rate-distortion problems. In this work, we study how the optimal privacy-utility tradeoff region is affected by constraints on the data that is directly available as input to the release mechanism. In particular, we consider the availability of only sensitive data, only useful data, and both (full data). We show that a general hierarchy holds, that is, the tradeoff region given only the sensitive data is no larger than the region given only the useful data, which in turn is clearly no larger than the region given both sensitive and useful data. In addition, we determine conditions under which the tradeoff region given only the useful data coincides with that given full data. This is based on the common information between the sensitive and useful data.**

I. Introduction

The objective of privacy-preserving data release is to provide useful data with minimal distortion while simultaneously minimizing the sensitive data revealed. Dependencies between the sensitive and useful data results in a privacy-utility tradeoff that has strong connections to generalized rate-distortion problems [1]. In this work, we study how the optimal privacy-utility tradeoff region is affected by constraints on the data that is directly available as input to the release mechanism. Such constraints are potentially motivated by applications where either the sensitive or useful data is not directly observable. For example, the useful data may be an unknown property that must be inferred from only the sensitive data. Alternatively, the constraints may be used to capture the limitations of a particular approach, such as *output-perturbation* data release mechanisms that take only the useful data as input, while ignoring the remaining sensitive data.

The general challenge of privacy-preserving data release has been the aim of a broad and varied field of study. Basic attempts to anonymize data have led to widely publicized leaks of sensitive information, such as [2], [3]. These have subsequently motivated a wide variety of statistical formulations and techniques for preserving privacy, such as k-anonymity [4], L-diversity [5], t-closeness [6], and differential privacy [7]. Our work concerns a non-asymptotic, information-theoretic treatment of this problem, such as in [1], [8], where the sensitive data and useful data are modeled as random variables X and Y, respectively, and mechanism design is the problem of constructing channels that obtain the optimal privacy-utility tradeoffs. While we consider a non-asymptotic, single-letter problem formulation, there are also related asymptotic coding problems that additionally consider communication efficiency in a rate-distortion-privacy tradeoff, as studied in [9], [10].

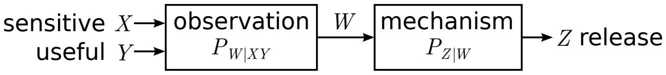

Fig. 1. The observation W of the sensitive data X and useful data Y is input to the data release mechanism which produces the released data Z.

In this work, we generalize the framework of [1], [8] to address scenarios with data constraints and allow for general utility metrics. In particular, we compare three scenarios, where only the sensitive data, only the useful data, or both (full data) are available. We show that a general hierarchy holds, that is, the tradeoff region given only the sensitive data is no larger than the region given only the useful data, which in turn is clearly no larger than the region given both sensitive and useful data. We also show that if the common information and mutual information between the sensitive and useful data are equal[1], then the tradeoff region given only the useful data coincides with that given full data, indicating when output perturbation is optimal despite unavailability of the sensitive data. Conversely, when the common information and mutual information are not equal, there exist distortion metrics where the tradeoff regions are not the same, indicating that output perturbation can be strictly suboptimal compared to the full data scenario.

II. Privacy-Utility Tradeoff Problem

Let X, Y, and W be discrete random variables (RVs) distributed on finite alphabets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{W}$, respectively. Let X denote the sensitive information that the user wishes to conceal, Y the useful information that the user is willing to reveal, and W the directly observable data, which may represent a noisy observation of X and/or Y. The target application dictates (or imposes a specific structure upon) the *data model* P_{XY} and *observation constraints* $P_{W|XY}$ so that $(X, Y, W) \sim P_{XY} P_{W|XY}$. The *data release mechanism* takes

[1] This statement applies for both the Wyner [11] and Gács-Körner [12] notions of common information.

W as input and (randomly) generates output Z in a given finite alphabet $\mathcal{Z}$ dictated by the target application (perhaps implicitly via the distortion metric). Note that Z must satisfy the Markov chain $(X, Y) \to W \to Z$ and the mechanism can be specified by the conditional distribution $P_{Z|W}$. A diagram of the overall system is shown in Figure 1.

The mechanism should be designed such that Z provides application-specific utility through the information it reveals about Y while protecting privacy by limiting the information it reveals about X.

A commonly used information-theoretic measure of privacy-leakage which quantifies the amount of information about X leaked by Z (on average) is the mutual information $I(X; Z)$ between them. We adopt this privacy-leakage measure in our work. *Privacy* is inversely related to $I(X; Z)$: privacy is stronger if the privacy-leakage $I(X; Z)$ is smaller. We have perfect privacy if $I(X; Z) = 0$. Thus, the aim is to minimize $I(X; Z)$ in order to maximize privacy.

The amount of *utility* that the mechanism-output Z provides about the useful information represented by Y can be quantified through a general distortion metric $D(P_{YZ})$, which is a functional that assigns values in $[0, \infty)$ to input joint distributions of Y and Z. Utility and distortion have an inverse relationship to each other: smaller the distortion, greater the utility. Thus, the aim is to minimize $D(P_{YZ})$. The specification of the distortion metric is dictated by the target application. Example distortion metrics include: 1) *expected distortion*, where $D(P_{YZ}) = E[d(Y, Z)]$ for some distortion function $d : \mathcal{Y} \times \mathcal{Z} \to [0, \infty)$, 2) *conditional entropy*, where $D(P_{YZ}) = H(Y|Z)$ which corresponds to the goal of maximizing the mutual information between Y and Z. Note that probability of error $\Pr(Y \neq Z)$ is an example within the class of expected distortion metrics where $d(y, z)$ is equal to zero when $y = z$ and equal to one otherwise.

Given a target application that specifies the data model P_{XY}, observation model $P_{W|XY}$, and distortion metric $D(P_{YZ})$, the goal of the system designer is to construct mechanisms $P_{Z|W}$ that provide the desired levels of privacy and utility while achieving the optimal tradeoff. We say that particular privacy-utility pair $(\epsilon, \delta) \in [0, \infty)^2$ is *achievable* if there exists a mechanism $P_{Z|W}$ with privacy leakage $I(X; Z) \leq \epsilon$ and distortion $D(P_{YZ}) \leq \delta$. The set of all achievable privacy-utility pairs forms the *achievable region* of privacy-utility tradeoffs. Particularly, we are interested the *optimal boundary* of this region, which can be expressed by the optimization problem

$$\begin{aligned} \pi(\delta) \triangleq \inf_{P_{Z|W}} & \ I(X; Z) \\ \text{s.t. } & D(P_{YZ}) \leq \delta, \end{aligned} \tag{1}$$

which determines the optimal privacy leakage as a function of the allowable distortion δ.

The distortion constraint, $D(P_{YZ}) \leq \delta$, can be equivalently expressed as a constraint on the conditional distribution $P_{Z|Y}$ since P_Y is fixed by the data model. Note that a mechanism specified by $P_{Z|W}$ determines the corresponding $P_{Z|Y}$ through the linear relationship[2]

$$P_{Z|Y}(z|y) = \sum_{w \in \mathcal{W}, x \in \mathcal{X}} P_{Z|W}(z|w) P_{W|XY}(w|x, y) P_{X|Y}(x|y). \tag{2}$$

Similarly, $P_{Z|X}$ is determined by $P_{Z|W}$ through the linear relationship

$$P_{Z|X}(z|x) = \sum_{w \in \mathcal{W}, y \in \mathcal{Y}} P_{Z|W}(z|w) P_{W|XY}(w|x, y) P_{Y|X}(y|x). \tag{3}$$

While general observation models $P_{W|XY}$ can be considered within this framework, particular structures may be of interest for certain applications. We highlight and explore the relationship between three specific cases for W, while allowing a general distribution P_{XY} between the sensitive and private data.

Full Data: In this case, P_{XY} is general but $W = (X, Y)$, capturing the situation when the mechanism has direct access to both the sensitive and useful information. For this case, the privacy-utility optimization problem of (1) reduces to

$$\begin{aligned} \pi_{\text{FD}}(\delta) \triangleq \inf_{P_{Z|XY}} & \ I(X; Z) \\ \text{s.t. } & D(P_{YZ}) \leq \delta. \end{aligned} \tag{4}$$

Output Perturbation: In this case, P_{XY} is general but $W = Y$, capturing the situation when the mechanism only has direct access to the useful information. For this case, the privacy-utility optimization problem of (1) reduces to

$$\begin{aligned} \pi_{\text{OP}}(\delta) \triangleq \inf_{P_{Z|Y}} & \ I(X; Z) \\ \text{s.t. } & D(P_{YZ}) \leq \delta, \end{aligned} \tag{5}$$

where $P_{Z|X}(z|x) = \sum_{y \in \mathcal{Y}} P_{Z|Y}(z|y) P_{Y|X}(y|x)$. Note: this optimization is equivalent to that of (4), with the Markov chain $X \to Y \to Z$ imposed as an additional constraint.

Inference: In this case, P_{XY} is general but $W = X$, capturing the situation when the mechanism only has direct access to the sensitive information and the useful information, such as a discrete hidden state, is not directly available or observable and needs to be *inferred* indirectly by processing the sensitive information. For this case, the privacy-utility optimization problem of (1) reduces to

$$\begin{aligned} \pi_{\text{INF}}(\delta) \triangleq \inf_{P_{Z|X}} & \ I(X; Z) \\ \text{s.t. } & D(P_{YZ}) \leq \delta, \end{aligned} \tag{6}$$

where $P_{Z|Y}(z|y) = \sum_{x \in \mathcal{X}} P_{Z|X}(z|x) P_{X|Y}(x|y)$. Note: this optimization is equivalent to that of (4), with the Markov chain $Y \to X \to Z$ imposed as an additional constraint.

III. Convexity and Rate-Distortion Connections

Here we discuss how for certain combinations of utility metrics and data constraints, the resulting tradeoff problem is equivalent to generalized rate-distortion and privacy-utility problems encountered in the literature. We also indicate how

[2] This and all other statements involving conditional distributions are defined only for symbols in the support of the conditioned random variables.

the tradeoff optimizations of (4), (5), and (6) will become convex for certain utility metrics.

Note that in the general tradeoff optimization problem (1), the distributions $P_{Z|X}$ and $P_{Z|Y}$ are *linear* functions of the optimization variable $P_{Z|W}$ as shown by (2) and (3), while P_{XYW} and its marginals are fixed. Thus, the convexity properties of the problem will follow from the convexity properties of the privacy and distortion metrics as functions of $P_{Z|X}$ and $P_{Z|Y}$, respectively. The mutual information privacy metric $I(X;Z)$ is a *convex* objective function of $P_{Z|X}$ and hence also of the optimization variable in each of the three scenarios given by (4), (5), and (6). Thus, for all convex distortion functionals, the overall optimization problem will be convex. For example, any expected distortion utility metric $D(P_{YZ}) = E[d(Y,Z)]$ is a linear and therefore a convex functional.

The privacy-utility tradeoff problem as considered by [1], [8] assumes the output perturbation constraint (see (5)), while using expected distortion $D(P_{YZ}) = E[d(Y,Z)]$ as the utility metric, and mutual information $I(X;Z)$ as the privacy metric. Additionally, [8] also considers maximum information leakage, $\max_{z \in \mathcal{Z}} [H(X) - H(X|Z=z)]$, as an alternative privacy metric. As noted by [8], the optimization problem for the full data scenario (see (4)) can be recast as an optimization with the output perturbation constraint, by redefining the useful data as $Y' := (X,Y)$ and the distortion function as $d'(Y',Z) := d(Y,Z)$. This approach allows one to solve the optimization problem for the full data scenario using an equivalent optimization problem appearing in the output perturbation scenario. However, the distinction between these two scenarios should not be overlooked, as the output perturbation scenario represents a fundamentally different problem where the sensitive data is not available, which in general results in a strictly smaller privacy-utility tradeoff region (see Theorem 3).

The inference scenario given by (6) with expected distortion $D(P_{YZ}) = E[d(Y,Z)]$ as the utility metric is equivalent to an indirect rate-distortion problem [13]. As shown by Witsenhausen in [13], indirect rate-distortion problems can be converted to direct ones with the modified distortion metric $d'(x,z) := E[d(Y,Z)|X=x, Z=z] = \sum_{y \in \mathcal{Y}} d(y,z) P_{Y|X}(y|x)$ since $Y \to X \to Z$ forms a Markov chain.

When the utility metric is conditional entropy, i.e., $D(P_{YZ}) = H(Y|Z)$, the equivalent utility objective is to maximize the mutual information $I(Y;Z)$, and the distortion constraint can be equivalently written as $I(Y;Z) \geq \delta'$, where $\delta' := H(Y) - \delta$. Thus, this results in the optimization problem of choosing Z to minimize $I(X;Z)$ subject to a lower bound on $I(Y;Z)$. This problem in the inference scenario, where the additional Markov chain constraint $Y \to X \to Z$ is imposed, is equivalent to the Information Bottleneck problem considered in [14], which also provides a generalization of the Blahut-Arimoto algorithm [15] to perform this optimization. For the output perturbation scenario, where the additional Markov chain constraint $X \to Y \to Z$ is imposed, this problem is called the Privacy Funnel and was proposed by [16]. In all three scenarios, the optimization problems are non-convex as the feasible regions are non-convex, specifically, they are complements of convex regions.

IV. Results

For a given (fixed) distribution P_{XY} between the sensitive and private data, we can study how the optimal privacy-utility tradeoff changes across the aforementioned three different cases of W. This is of practical interest, since the restrictions on W in the inference and output perturbation mechanisms might be considered not just for when these situations inherently arise in the given application, but also for simplifying mechanism design and optimization.

Since the optimization problems of (5) and (6) are equivalent to (4) with an additional Markov chain constraint, we immediately have that $\pi_{\text{FD}}(\delta) \leq \pi_{\text{OP}}(\delta)$ and $\pi_{\text{FD}}(\delta) \leq \pi_{\text{INF}}(\delta)$ for any δ. This implies that the achievable privacy-utility regions of both the inference mechanism and output perturbation mechanism are contained within the achievable privacy-utility region of the full data mechanism, which intuitively follows since the full data mechanism only has more input data available. The next theorem establishes the general relationship between the inference and output perturbation tradeoff regions.

Theorem 1. ***(Output Perturbation better than Inference)*** *For any data model P_{XY} and distortion metric $D(P_{YZ})$, the achievable privacy-utility region for the output perturbation mechanism (when $W = Y$) contains the achievable privacy-utility region for the inference mechanism (when $W = X$), that is, $\pi_{OP}(\delta) \leq \pi_{INF}(\delta)$ for any δ.*

Combining the preceding theorem with the earlier observations, we have that $\pi_{\text{FD}}(\delta) \leq \pi_{\text{OP}}(\delta) \leq \pi_{\text{INF}}(\delta)$ for any δ. Thus, in general, full data offers a better privacy-utility tradeoff than output perturbation, which in turn offers a better privacy-utility tradeoff than inference.

The next theorem establishes that for a certain class of joint distributions P_{XY}, the full data and output perturbation mechanisms achieve the same optimal privacy-utility tradeoff. Thus, for this class of P_{XY}, the full data mechanism design can be simplified to the design of an output perturbation mechanism, which can ignore the sensitive data X without degrading the privacy-utility performance. Specifically, this class is characterized by those joint distributions P_{XY} for which common information $C(X;Y) = I(X;Y)$. Some of the key properties of common information that are needed for proving Theorems 2 and 3 are summarized in Appendix A.

Theorem 2. ***(Sufficient Conditions for the General Optimality of Output Perturbation)*** *For any distortion metric $D(P_{YZ})$ and any data model P_{XY} where $C(X;Y) = I(X;Y)$, the achievable privacy-utility region for the output perturbation mechanism (when $W = Y$) is the same as the achievable privacy-utility region for the full data mechanism (when $W = (X,Y)$), that is, $\pi_{OP}(\delta) = \pi_{FD}(\delta)$ for any distortion metric and any δ.*

Theorem 2 establishes that $C(X;Y) = I(X;Y)$ is a sufficient condition on P_{XY} such that, for any general distortion metric, full data mechanisms cannot provide better privacy-utility tradeoffs than the output perturbation mechanisms. Our next theorem gives the converse result, establishing that for data models where $C(X;Y) \neq I(X;Y)$, output perturbation mechanisms are generally suboptimal, that is, there exists a distortion metric such that the full data mechanisms provide a strictly better privacy-utility tradeoff.

Theorem 3. ***(Necessary Conditions for the General Optimality of Output Perturbation)*** *For any data model P_{XY} where $C(X;Y) \neq I(X;Y)$, there exists a distortion metric $D(P_{YZ})$ such that the achievable privacy-utility region for the output perturbation mechanism (when $W = Y$) is strictly smaller than the achievable privacy-utility region for the full data mechanism (when $W = (X,Y)$), that is, there exists $\delta \geq 0$ such that $\pi_{OP}(\delta) > \pi_{FD}(\delta)$.*

V. Conclusion

In this paper, we formulated the privacy-utility tradeoff problem where the data release mechanism has limited access to the entire data composed of useful and sensitive parts. Based on the information theoretic formulation, we compared the privacy-utility tradeoff regions attained by full data, output perturbation, and inference mechanisms, which have access to the entire data, only useful data, and only sensitive data, respectively.

We first observed that the full data mechanism provides the best privacy-utility tradeoff and then showed that the output perturbation mechanism provides a better privacy-utility tradeoff than the inference mechanism. We showed that if the common and mutual information between useful and sensitive data are identical, then the full data mechanism simplifies to the output perturbation mechanism. Conversely, we showed that if the common information is not equal to mutual information, then the tradeoff region achieved by full data mechanism is strictly larger than the one achieved by the output perturbation mechanism.

Throughout the paper, we allowed for a general distortion metric, but focused specifically on mutual information as the privacy metric. In our ongoing work, we are investigating the privacy-utility tradeoff problem for general privacy metrics and observation models, and the evaluation of tradeoff regions attainable for non-trivial data models.

References

[1] D. Rebollo-Monedero, J. Forné, and J. Domingo-Ferrer, "From t-closeness-like privacy to postrandomization via information theory," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 11, pp. 1623–1636, 2010.

[2] L. Sweeney, "Simple demographics often identify people uniquely," *Carnegie Mellon University, Data Privacy Working Paper*, 2000.

[3] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *IEEE Symp. on Security and Privacy*. IEEE, 2008, pp. 111–125.

[4] L. Sweeney, "k-anonymity: A model for protecting privacy," *Intl. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[5] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," *ACM Trans. on Knowledge Discovery from Data*, vol. 1, no. 1, p. 3, 2007.

[6] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *IEEE Intl. Conf. on Data Eng.* IEEE, 2007, pp. 106–115.

[7] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography*. Springer, 2006, pp. 265–284.

[8] F. du Pin Calmon and N. Fawaz, "Privacy against statistical inference," in *Allerton Conf. on Comm., Ctrl., and Comp.*, 2012, pp. 1401–1408.

[9] H. Yamamoto, "A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers," *IEEE Trans. on Information Theory*, vol. 29, no. 6, pp. 918–923, 1983.

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

[11] A. Wyner, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 163–179, Mar. 1975.

[12] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 149–162, 1973.

[13] H. S. Witsenhausen, "Indirect rate distortion problems," *IEEE Transactions on Information Theory*, vol. 26, no. 5, pp. 518–521, 1980.

[14] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *Allerton Conf. on Comm., Ctrl., and Comp.*, 1999, pp. 368–377.

[15] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. John Wiley & Sons, 2012.

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

[17] R. Ahlswede and J. Körner, "On common information and related characteristics of correlated information sources," in *Proc. Prague Conf. on Information Theory*, 1974.

Appendix A
Properties of Common Information

The *graphical representation* of P_{XY} is the bipartite graph with an edge between $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ if and only if $P_{XY}(x,y) > 0$. The *common part* U of two random variables (X,Y) is defined as the (unique) label of the connected component of the graphical representation of P_{XY} in which (X,Y) falls. Note that U is a deterministic function of X alone and also a deterministic function of Y alone.

The Gács-Körner common information of two random variables (X,Y) is given by entropy of the common part, that is, $C(X;Y) := H(U)$, and has the operational significance of being the maximum number of common bits per symbol that can be independently extracted from X and Y [12]. In general, $C(X;Y) \leq I(X;Y)$, with equality if and only if $X \to U \to Y$ forms a Markov chain [17]. Since our results are only concerned with whether $C(X;Y) = I(X;Y)$, our theorem statements are unchanged if we use instead the Wyner notion of common information (see [11]), since it is also equal to mutual information if and only if $X \to U \to Y$ forms a Markov chain [17].

We give the following lemma which aids our proof of Theorem 3 in Appendix D.

Lemma 1. *If $C(X;Y) \neq I(X;Y)$, then there exist $x_0, x_1 \in \mathcal{X}$ and $y_0, y_1 \in \mathcal{Y}$, such that $y_0 \neq y_1$, $P_{XY}(x_0,y_0) > 0$, $P_{XY}(x_0,y_1) > 0$, and $P_{X|Y}(x_1|y_0) \neq P_{X|Y}(x_1|y_1)$.*

Proof: We will prove this lemma by showing the contrapositive, that is, if there does not exist $x_0, x_1 \in \mathcal{X}$ and $y_0, y_1 \in \mathcal{Y}$ satisfying the conditions stated in the lemma, then $C(X;Y) = I(X;Y)$. First, note that if for all $x_0 \in \mathcal{X}$ and $y_0, y_1 \in \mathcal{Y}$, either $y_0 = y_1$, $P_{XY}(x_0, y_0) = 0$, or $P_{XY}(x_0, y_1) = 0$, then Y is a deterministic function of X, which would result in $C(X;Y) = I(X;Y)$. Thus, we are left with showing that for all $x_0 \in \mathcal{X}$ and $y_0, y_1 \in \mathcal{Y}$, with $y_0 \neq y_1$, $P_{XY}(x_0, y_0) > 0$, and $P_{XY}(x_0, y_1) > 0$, if we also have that for all $x_1 \in \mathcal{X}$, $P_{X|Y}(x_1|y_0) = P_{X|Y}(x_1|y_1)$, then $C(X;Y) = I(X;Y)$. This follows since these conditions would imply that for the common part U of (X,Y), $X \to U \to Y$ forms a Markov chain. ■

Appendix B
Proof of Theorem 1

It is sufficient to show that for any mechanism $P_{Z|X}$ that is a feasible solution in the inference optimization of (6), there is a corresponding mechanism $P_{Z'|Y}$ for the output perturbation optimization of (5) that achieves the same distortion and only lesser or equal privacy-leakage.

Let $P_{Z|X}$ be a mechanism in the feasible region of the inference optimization problem of (6). Define the corresponding mechanism for the output perturbation optimization of (5) by

$$P_{Z'|Y}(z|y) := \sum_{x \in \mathcal{X}} P_{Z|X}(z|x) P_{X|Y}(x|y).$$

Let $(X, Y, Z, Z') \sim P_{XY} P_{Z|X} P_{Z'|Y}$. Note that by construction, (Y, Z) and (Y, Z') have the same distribution $P_Y P_{Z'|Y}$. Thus, both mechanisms achieve the same distortion $D(P_Y P_{Z'|Y})$ and $I(Y;Z) = I(Y;Z')$. Further, by construction, $Y \to X \to Z$ and $X \to Y \to Z'$ form Markov chains. Thus, by the data processing inequality,

$$I(X;Z') \leq I(Y;Z') = I(Y;Z) \leq I(X;Z),$$

showing that the output perturbation mechanism has only lesser or equal privacy-leakage.

Appendix C
Proof of Theorem 2

Since $\pi_{\text{FD}}(\delta) \leq \pi_{\text{OP}}(\delta)$ is immediate, we only need to show that $\pi_{\text{OP}}(\delta) \leq \pi_{\text{FD}}(\delta)$. It is sufficient to show that for any mechanism $P_{Z|XY}$ that is a feasible solution in the full data optimization of (4), there is a corresponding mechanism $P_{Z'|Y}$ for the output perturbation optimization of (5) that achieves the same distortion and only lesser or equal privacy-leakage.

Let $P_{Z|XY}$ be a mechanism in the feasible region of the full data optimization problem of (4). Define the corresponding mechanism for the output perturbation optimization of (5) by

$$P_{Z'|Y}(z|y) := \sum_{x \in \mathcal{X}} P_{Z|XY}(z|x, y) P_{X|Y}(x|y).$$

Let $(X, Y, Z, Z') \sim P_{XY} P_{Z|XY} P_{Z'|Y}$, and let U be the common part of (X, Y), where, by construction, U is a deterministic function of either X alone or Y alone. Since $C(X;Y) = I(X;Y)$, we have that $X \to U \to Y$ forms a Markov chain, i.e., $I(X;Y|U) = 0$. By construction, $X \to Y \to Z'$ also forms a Markov chain, and hence $I(X;Z'|UY) = I(X;Z'|Y) = 0$, since U is deterministic function of Y. Given these two Markov chains, we have

$$\begin{aligned} 0 &= I(X;Y|U) + I(X;Z'|UY) \\ &= I(X;YZ'|U) \\ &= I(X;Z'|U) + I(X;Y|UZ') \\ &\geq I(X;Z'|U), \end{aligned}$$

and hence $I(X;Z'|U) = 0$, i.e., $X \to U \to Z'$ also forms a Markov chain. Continuing, we can show the desired privacy-leakage inequality,

$$\begin{aligned} I(X;Z') &\overset{(a)}{=} I(XU;Z') \\ &= I(U;Z') + I(X;Z'|U) \\ &\overset{(b)}{=} I(U;Z) \\ &\leq I(U;Z) + I(X;Z|U) \\ &= I(XU;Z) \\ &\overset{(c)}{=} I(X;Z), \end{aligned}$$

where (a) and (c) follow from U being a deterministic function of X, and (b) follows from the fact that $P_{YZ} = P_{YZ'}$ (and hence $P_{UZ} = P_{UZ'}$) by construction and the Markov chain $X \to U \to Z'$.

Appendix D
Proof of Theorem 3

We will show the following result, which is key to the proof.

Lemma 2. *If $C(X;Y) \neq I(X;Y)$ then there exist random variables Z and Z' with $P_{YZ} = P_{YZ'}$, such that $X \to Y \to Z'$ forms a Markov chain, $I(X;Z) = 0$, and $I(X;Z') > 0$.*

The proof of Theorem 3 then follows by defining the distortion functional (metric) $D(P_{YZ})$ to equal 1 for the particular choice of $P_{YZ'}$ in Lemma 2 and to equal 2 otherwise, and choosing $\delta = 1$. This choice for the distortion metric and distortion level restricts the feasible output perturbation mechanism to only $P_{Z'|Y}$, which by Lemma 2 results in $\pi_{OP}(\delta) = I(X;Z') > 0$. However, Lemma 2 also ensures the existence of Z produced by a full data mechanism $P_{Z|XY}$ that results in $\pi_{FD}(\delta) = I(X;Z) = 0$.

Using the symbols (x_0, x_1, y_0, y_1) shown to exist by Lemma 1, we can prove Lemma 2 by constructing a binary Z with alphabet $\mathcal{Z} = \{0, 1\}$ as follows. Choose any $s \in (0,1)$ and any $t \in \big(0, \min\{s'/P_{Y|X}(y_1|x_0), s/P_{Y|X}(y_0|x_0)\}\big)$, where $s' := (1 - s)$. Define Z with $(X, Y, Z) \sim P_{XY} P_{Z|XY}$, where

$$P_{Z|XY}(0|x, y) := \begin{cases} s + tP_{Y|X}(y_1|x_0), & \text{if } (x, y) = (x_0, y_0), \\ s - tP_{Y|X}(y_0|x_0), & \text{if } (x, y) = (x_0, y_1), \\ s, & \text{otherwise.} \end{cases}$$

The choice of s and t ensures that $P_{Z|XY}(0|x, y) \in (0, 1)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. This construction of $P_{Z|XY}$ makes Z independent of X, since for all $x \in \mathcal{X}$ in the support of P_X,

$$P_{Z|X}(0|x) = \sum_{y \in \mathcal{Y}} P_{Z|XY}(0|x, y) P_{Y|X}(y|x) = s.$$

With the above construction, we have

$$
\begin{aligned}
P_{Z|Y}(0|y) &= \sum_{x\in\mathcal{X}} P_{Z|XY}(0|x,y)P_{X|Y}(x|y)\\
&= \begin{cases} s + tP_{Y|X}(y_1|x_0)P_{X|Y}(x_0|y_0), & \text{if } y = y_0,\\ s - tP_{Y|X}(y_0|x_0)P_{X|Y}(x_0|y_1), & \text{if } y = y_1,\\ s, & \text{otherwise.}\end{cases}
\end{aligned}
$$

Next, we construct binary Z' such that $X \to Y \to Z'$ forms a Markov chain, with $(X, Y, Z') \sim P_{XY}P_{Z'|Y}$, where we set $P_{Z'|Y} := P_{Z|Y}$. Then, consider

$$
\begin{aligned}
P_{Z'|X}(0|x) &= \sum_{y\in\mathcal{Y}} P_{Z'|Y}(0|y)P_{Y|X}(y|x)\\
&= \sum_{y\in\mathcal{Y}} P_{Z|Y}(0|y)P_{Y|X}(y|x)\\
&= s + tP_{Y|X}(y_1|x_0)P_{X|Y}(x_0|y_0)P_{Y|X}(y_0|x)\\
&\quad - tP_{Y|X}(y_0|x_0)P_{X|Y}(x_0|y_1)P_{Y|X}(y_1|x)\\
&= s + tP_X(x_0)P_{Y|X}(y_0|x_0)P_{Y|X}(y_1|x_0)\\
&\quad\times [P_{X|Y}(x|y_0) - P_{X|Y}(x|y_1)]/P_X(x).
\end{aligned}
$$

Finally, we show that $P_{Z'|X}(0|x)$ is not constant for all $x \in \mathcal{X}$ in the support of P_X, which implies that Z' is not independent of X, i.e., $I(X; Z') > 0$. This can be proved by contradiction, by supposing that $P_{Z'|X}(0|x)$ is constant for all $x \in \mathcal{X}$ in the support of P_X. Then, for all $x \in \mathcal{X}$,

$$P_{X|Y}(x|y_0) - P_{X|Y}(x|y_1) = cP_X(x),$$

for some constant c. By summing over all $x \in \mathcal{X}$, we have that $c = 0$. This would imply that $P_{X|Y}(x|y_0) = P_{X|Y}(x|y_1)$ for all $x \in \mathcal{X}$, contradicting the existence of $x_1 \in \mathcal{X}$ given by Lemma 1 for the choice of y_0 and y_1.