Computing Messages that Reveal Selected Inferences While Protecting Others

Matthew P. Johnson and Liang Zhao
Lehman College and the Graduate Center, City University of New York

Supriyo Chakraborty
IBM Thomas J. Watson Research Center

***Abstract*—We study a lossy coding scenario, posed as an algorithmic optimization problem, where we trade off between the conflicting goals of accuracy and privacy, motivated by scenarios such as the public release of *estimates* of data that accurately reflect *some* aspects of the raw data without revealing other sensitive confidential aspects of it, or permitting it to be inferred. More precisely, given a discrete probability distribution $p(D, X, Y)$, where X represents the *whitelisted* inferences from D and Y represents the *blacklisted* inferences, we seek to construct a conditional distribution $p(M|D)$ with the dual goals of making $I(M; X)$ large and $I(M; Y)$ small. Chakraborty et al. 2013 [1] provided optimal solutions within this model to the two extreme points on the two objectives' Pareto frontier: maximizing $I(M; X)$ subject to the constraint that $I(M; Y)$ be as small as possible ("perfect privacy", using linear programming (LP)) and vice versa ("perfect utility", which is trivial).**

In this paper we provide a faster *combinatorial* optimal algorithm for the perfect privacy problem, which does not require the use of an LP solver. Moreover, this algorithm any be used to compute Pareto-optimal solutions at *any* point on the Pareto frontier. (En route to this algorithm, we also provide a mathematical programming-based solution.) This solves the primary open problem posed by [1].

I. Introduction

In a wide variety of applications, (perturbed) data is shared subject to conflicting objectives involving accuracy and privacy: the provider derives utility of some kind from the resulting actions or information-processing performed by the recipient, but at the same time the provider may, for privacy reasons, wish to shield certain sensitive portions of the data or prevent the receiver from drawing certain valid but undesirable inferences from the data. For example, a medical application uses data from a respiration sensor to infer breathing irregularities (utility), but the same data may also reveal smoking habits or the onset of stress (unwanted inference) [2]; similarly, data from an accelerometer is typically used by an application for detecting phone orientation or user activity (utility), but the same data could also be used to infer keystrokes, or user speech (unwanted inferences) [3], [4], [5].

We study this problem within the framework provided by Chakraborty et al. [1], which gave an abstract information-theoretic model of such privacy/utility tradeoff scenarios, posing them as a bicriteria optimization problem. Given is a discrete probability distribution $p(D, X, Y)$, where D is a random variable representing the underlying data, X represents the class of *whitelisted* utility-providing inferences we want the recipient to be able to draw from the message M, and Y represents the *blacklisted* sensitive inferences we want to keep private, the task is to construct a conditional distribution $p(M|D)$ with the dual goals of making $I(M; X)$ large and $I(M; Y)$ small.

Prior work focused on the two extreme points of this tradeoff: 1) optimizing utility under a constraint of "perfect" privacy ($\max_{\text{perfP}} \mathrm{I(M;X)}$), and 2) optimizing privacy under a constraint of "perfect" utility ($\min_{\text{perfU}} \mathrm{I(M;Y)}$). In this paper we connect the gap between these two extreme points, considering all possible intermediate privacy/utility tradeoffs, i.e., solutions to the problem of minimizing $I(M; Y) - \beta I(M; X)$ with all possible weights $\beta \geq 0$. Equivalently, these are all possible Pareto-optimal solutions to the bicriteria problem (i.e., solutions where improving $I(M; X)$ would require degrading $I(M; Y)$, and vice versa), which collectively form the Pareto frontier. (Note that the extreme points above are also included in this frontier.)

Contributions. Our main result is a combinatorial algorithm which optimally solves the $\max_{\text{perfP}} \mathrm{I(M;X)}$ faster than the linear programming-based solution of [1] and moreover can compute Pareto-optimal solutions corresponding to *all* possible privacy/utility tradeoffs on this problem's Pareto frontier. (We also give a mathematical programming-based solution.) This solves the primary open problem posed by [1].

Related work. Our investigation, while motivated by the approaches in [6], [7], [8], [9], differs from their settings in several ways. First, our scheme does not depend on the existence of a multi-user database as in [6]. Second, unlike in [8], [9], the private variable Y and non-private variable X are (in our model) arbitrary deterministic functions of the same data D. Finally, in our setting, unlike in the Information Bottleneck (IB) method [7], what the provider wishes to keep private is some arbitrary function, *different from identity*, computed on the raw data and not the raw data itself. Thus, our goal of minimizing the mutual information $I(M; Y)$ between M and a *specific function* $Y = g(D)$ on the input data generalizes the IB method that minimizes $I(M; D)$.

Organization. The rest of this paper is organized as follows. After related work, we review and extend the system model of Chakraborty et al. [1] in Sect. II. We sketch the proofs of a number of properties charactering Pareto-optimal solutions to our problem in Sect. III. (Due to space limitations, most of the proofs are deferred to the full version of the paper.) We give a combinatorial algorithm for this problem in Sect. IV. We conclude in Sect. V.

II. Preliminaries

In this model, a data provider possesses a discrete random variable (rv) D. The provider wishes to share information about the *whitelist* specified by the rv $X = f(D)$, but wants to keep private the *blacklist* specified by rv $Y = g(D)$ private. X and Y are assumed, *without loss of generality*, to be deterministic functions of D. In this work, we focus on strategies for generating the message M from D, specified by the distribution $p(M|D)$ achieving a desired tradeoff between the goals of utility (M gives a lot of information about X) and privacy (M reveals only limited information about Y). Since $p(D)$ is given, we can equivalently speak in terms of computing the joint distribution $p(M, D)$.

Notation. Throughout the paper we will consider single-letter characterization of D, X, Y and M. The (finite) alphabets of D, X, Y, M are denoted by $\mathcal{D}$, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{M}$, respectively, and individual members of them by d, x, y, m. When one of these lower-case member variables appears in a summation or set notation, it is understood to range over the entire corresponding alphabet unless otherwise restricted.

Formally, for a given choice of distribution $p(M|D)$ of sending message M given data D, we incur a utility penalty $\delta_U(M)$ and a privacy penalty $\delta_P(M)$:

$$\begin{aligned} \delta_U(M) &\triangleq \frac{H(X|M)}{H(X)} = 1 - \frac{I(X;M)}{H(X)} \\ \delta_P(M) &\triangleq \frac{I(Y;M)}{H(Y)} = 1 - \frac{H(Y|M)}{H(Y)} \end{aligned}$$

Definition 1. *We define the* support of m *as the set* $S(m) = \{d : p(d|m) > 0\}$ *and the* support of d *as the set* $S(d) = \{m : p(d|m) > 0\}$. *Similarly,* $S(m, x) = \{y : p(x, y|m) > 0\}$ *and* $S(m, x) = \{x : p(x, y|m) > 0\}$. *We say* (x, y, m) appear together *if* $p(x, y|m) > 0$, *and similarly for* (x, m) *and* (y, m); *we say* (x, y) appear together *if* $p(x, y) > 0$.

We state the definitions involving m in terms of conditional probabilities rather than joint probabilities in order that a (potential) message m can be unambiguously identified with its support even if it is not used, i.e., if $p(m) = 0$.

We now state the first of several normalization assumptions that will be used to limit the search space of messages considered for use in Pareto-optimal solutions.

Condition 1. $d_1 = d_2$ *iff* $x_1 = x_2$ *and* $y_1 = y_2$ *(where* $(x_i, y_i) = (f(d_i), g(d_i))$*).*

Lemma 1. *Without loss of generality, any* $p(D, X, Y)$ *and* $p(M)$ *can be assumed to satisfy Cond. 1.*

Proof: If $f(d_1) = f(d_2)$ and $g(d_1) = g(d_2)$, then d_1 and d_2 can be merged into a single d without affecting δ_U or δ_P; conversely, if a single d appears with multiple pairs $(x_1, y_1), (x_2, y_2)$, then d can be replaced by new symbols d_1 and d_2, updating the probabilities so that $p(d_i, x_i, y_i) = p(d, x_i, y_i)$ for $i = 1, 2$. □

Identifications. Thus observe w.l.o.g. that each value d can be identified with (i.e., is uniquely specified by) the pair $(x, y) = (f(d), g(d))$, and hence specifying $p(X, Y)$ is equivalent to specifying $p(D)$, assuming that $f(\cdot)$ and $g(\cdot)$ are fixed. Throughout the paper we will treat values d as pairs (x, y) when convenient. That is, we are free to treat d as either the fundamental entity, with x and y being functions of it, or alternatively treat the pair (x, y) as the fundamental entity, with d simply its *name* and with $f(d)$ and $g(d)$ as *projections* from (x, y) (in the relational databases sense). Thus the problem instance can be specified equivalently as either $p(D)$ or $p(X, Y)$. We will prove below that each message m can be identified with its support $S(m)$, i.e., with some subset of $\mathcal{D}$, w.l.o.g., in the sense that there is no added penalty for restricting our attention to such solutions.

Tradeoff between objectives. The parameters δ_P, δ_U expressed in terms of information-theoretic notions conveniently capture the tradeoff between privacy and utility. The smaller $\delta_U(M)$ is, the more useful M is in determining X, and the smaller $\delta_P(M)$ is, the more private Y remains when M is revealed. We refer to $\delta_U = 0$ as the *perfect utility* case (in which $H(X|M) = 0$ and therefore X can be perfectly inferred from M) and to $\delta_P = 0$ as the *perfect privacy* case (in which $I(Y; M) = 0$, i.e., $H(Y|M) = H(Y)$, and therefore Y is independent of M). Analogous to false-alarm/missed-detection tradeoffs [10], in general (i.e., whenever $I(X; Y) > 0$) it is not possible to achieve perfect utility and perfect privacy at the same time (Fig. 1b), while optimizing for either on its own would be trivial. The two extremes in terms of how we can trade off between the two goals are optimizing δ_U subject to the constraint that $\delta_P = 0$ ($\min_{\text{perfP}} \delta_U$) and vice versa ($\min_{\text{perfU}} \delta_P$).

More generally, we can seek to optimize one of the objectives given a bound c on the value of the other, i.e., $\min\{\delta_P(M) : \delta_U(M) \leq c\}$ or $\min\{\delta_U(M) : \delta_P(M) \leq c\}$. Note that these two problems are equivalent in the sense that if we can solve one of them then we can solve the other to arbitrary precision through binary search. A third equivalent formulation of the problem is to minimize a weighted average of the two objectives (for some constant $\beta \geq 0$):

$$\mathcal{L}[p(m|d)] = \delta_P(M) + \beta\delta_U(M)$$

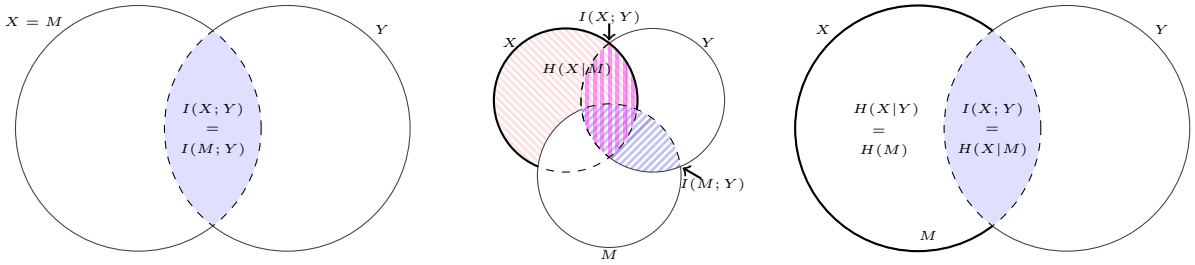
or equivalently (after appropriate normalization; we will alternate between these two formulations as convenient):

$$\mathcal{L}[p(m|d)] \sim I(M; Y) - \beta I(M; X)$$

III. Characterizing Optimal Solutions

Chakraborty et al. [1] obtained a linear programming-based solution to $\min_{\text{perfU}} \delta_{\text{P}}$ by proving a series of necessary conditions on its optimal solutions, which permitted them to restrict their attention to messages corresponding to a columns in a predefined table. First they prove that M and Y are independent, which holds for optimal $\min_{\text{perfP}} \delta_{\text{U}}$ solutions, although for Pareto-optimal solutions in general.

Then (via an intermediate property) they show that each (m, y) pair will be associated with *exactly one* x, meaning that each message in an optimal $\min_{\text{perfP}} \delta_{\text{U}}$ solution will appear with exactly $|\mathcal{Y}|$ of the inputs $\mathcal{X} \times \mathcal{Y}$. For Pareto-optimal solutions we must relax this, replacing "exactly one" with "at most one": it is now possible that $p(y|m) = 0$, with m appearing with



(a) OptP PerfU. (b) General situation. (c) OptU PerfP (*ideal*).

Fig. 1: Illustration of the relationships between random variables X, Y, M under different scenarios: PerfU (upper left points in Fig. 2 below), intermediate points on the Pareto frontier, and PerfP (lower right points in Fig. 2; typically the optimal solution will have $H(X|M) > I(X;Y)$).

fewer than $|\mathcal{Y}|$ inputs (see the example solutions shown in Fig. 2).

Condition 2. *For each pair (m, y), there is at most one x appearing with (m, y), i.e., $|S(m, y)| \leq 1$. Also, w.l.o.g. $|S(m)| \geq 1$ (to exclude messages with null support).*

Proving that this relaxed condition holds for Pareto-optimal solutions turns out to be quite nontrivial (proof omitted).

Another assumption implicitly relied on by [1] is that no two distinct messages will have the same support.

Condition 3. $S(m_1) = S(m_2)$ *implies* $m_1 = m_2$.

Now, unlike those above this is not a *necessary* condition, since e.g. splitting the weight of a column in a $\min_{\text{perfP}} \delta_{\text{U}}$ solution equally between two new identical columns would not affect the solution value, but it is easy to show for *optimal* $\min_{\text{perfP}} \delta_{\text{U}}$ *solutions* that this very natural assumption can be made without loss of generality in the sense that there always exist optimal solutions satisfying this condition.

This continues to hold for Pareto-optimal solutions, but proving it turns out to be quite nontrivial (proof omitted).

Let $\mathcal{S}$ be the set of all possible supports consistent with Conds. 1-3. The choice that a given support makes concerning y is to either choose one $d \in g^{-1}(y)$ or else choose no such d. Each support can be thought of as one combination of such choices for all y, which implies that $|\mathcal{S}| = \prod_{y \in \mathcal{Y}} \left| g^{-1}(y) + 1 \right| - 1$, where $|\mathcal{M}| \leq |\mathcal{S}|$ (see Fig. 2).

IV. Approximating the Pareto Frontier

In this section we obtain an iterative algorithm for our problem, inspired by the iterative algorithm for the IB problem. Ours involves four rather than three random variables, which significantly complicates the analysis, though the resulting algorithm is a relatively simple iterative algorithm that repeatedly alternates between recomputing the probabilities $p(m|y)$ and those of related distributions, initialized with arbitrary values (see Algorithm 1). The algorithm's running time is $O(i \cdot |\mathcal{M}|)$, where i is the number of iterations.

We begin by computing the derivative of $\mathcal{L}$ with respect to $p(m|x, y)$ (for each triple (m, x, y)) and setting the result to zero (i.e., $\frac{\partial \mathcal{L}}{\partial p(d|m)} = 0$), adapting and extending the analysis of [7]. This allows us to compute an equation for $p(m|x, y)$ which together with a set of self-consistent equations (defined below) will lead to an iterative algorithm for computing the values $p(m|x, y)$.

We know we can restrict ourselves to solutions satisfying $|S(m, y)| \leq 1, \forall y$ (Cond. 2). For any tuple (m, y) satisfying $S(m, y) = 1$, we use $x_{m,y}$ to denote the unique element in $S(m, y)$. Via the following deduction:

$$\begin{aligned} p(m|x, y) &= p(m, x, y)/p(x, y) \\ &= p(x|m, y)p(m|y)p(y)/p(x, y) \\ &= \begin{cases} p(m|y)/p(x|y) & \text{if } x = x_{m,y} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{C2}$$

we can express the objective function as a functional depending *only* on distribution $p(M|Y)$, which determines the distributions $p(M)$ and $p(X|M)$:[1]

$$\begin{aligned} \mathcal{F}^{M|Y} &:= I(M;Y) - \beta I(M;X) \\ &= \sum_{m,y} p(m|y)p(y) \left(\log \frac{p(m|y)}{p(m)} - \beta \log \frac{p(x_{m,y}|m)}{p(x_{m,y})} \right) \end{aligned} \tag{$\mathcal{F}^{M|Y}$}$$

In addition, let $\mathcal{F}^3$ indicate the functional that is identical to $\mathcal{F}^{M|Y}$, except that it takes *three* sets of parameters $p(m|y)$, $p(m)$, and $p(x|m)$, which are permitted to vary independently. $\mathcal{F}^3$ will be used in the analysis below.

We know that we can further restrict ourselves to no duplicate supports (Cond. 3). Thus m is understood to range over the set $\mathcal{M}$ throughout. Therefore we formulate the problem as the following mathematical program, which turns out to be a convex optimization problem:

Problem 1

(Inst: $p(X, Y)$, Sol: $p(M|Y)$)

$$\begin{aligned} &\min \mathcal{F}^{M|Y} = I(M;Y) - \beta I(M;X) \\ &\text{s.t.} \sum_{m \in S(x,y)} p(m|y) = p(x|y), \quad \forall\ x, y \end{aligned} \tag{P1}$$

We can show (proof omitted) the following.

[1] To simplify notation we let capital-letter inputs to $p(\cdot)$ denote an entire discrete distribution, i.e., a set of probability values ranging over all possible inputs, e.g., $p(M|Y) = \{p(m|y)\}_{m,y}$, $p(M) = \{p(m)\}_m$, $p(X|M) = \{p(x|m)\}_{m,x}$.

Theorem 1. *A solution to **P1**, i.e., a conditional probability distribution $p(M|Y)$, is a stationary point of $\mathcal{F}^{M|Y} = I(M;Y) - \beta I(M;X)$ if and only if*

$$p(m|y) = \frac{1}{Z(x_{m,y},y)}\left(p(m)\left(\frac{p(x_{m,y}|m)}{p(x_{m,y})}\right)^{\beta}\right), \quad \forall\ m,y, \quad (1)$$

where $Z(x,y) = \frac{1}{p(x|y)} \cdot \sum_{m\in S(x,y)} \left(p(m)\left(\frac{p(x|m)}{p(x)}\right)^{\beta}\right)$.

Unfortunately Eq. (1) does not provide an explicit solution to **P1**, as it relies on unknown quantities $p(m)$ and $p(x_{m,y}|m)$. Instead, we analyze the three sets of unknown quantities ($\{p(M|Y)\}$, $\{p(M)\}$, $\{p(X|M)\}$) separately, through three subproblems. In each subproblem, one of the three sets of unknowns comprises the variables being optimized over, and the other two sets of unknowns are taken as constants which (together with $p(X,Y)$) specify the subproblem instance. Repeatedly solving these subproblems in terms of the others' previous solutions yields an iterative algorithm.

Subproblem 2a

(Inst: $(p(X,Y), p(M), p(X|M))$, Sol: $p(M|Y)$)

$$\min I(M;Y) - \beta I(M;X) \quad \textbf{(P2a)}$$
$$\text{s.t.} \sum_{m\in S(x,y)} p(m|y) = p(x|y), \quad \forall\ x,y$$

Subproblem 2b

(Inst: $(p(X,Y), p(M|Y), p(X|M))$, Sol: $p(M)$)

$$\min I(M;Y) - \beta I(M;X) \quad \textbf{(P2b)}$$
$$\text{s.t.} \sum_{m} p(m) = 1$$

Subproblem 2c

(Inst: $(p(X,Y), p(M|Y), p(M))$, Sol: $p(X|M)$)

$$\min I(M;Y) - \beta I(M;X) \quad \textbf{(P2c)}$$
$$\text{s.t.} \sum_{x} p(x|m) = 1, \quad \forall\ m$$

Definition 2. *We say $(p(M|Y), p(M), p(X|M))$ are* self-consistent *if $p(m) = \sum_y p(m|y)p(y)$, $\forall m$ and $p(x|m) = \frac{\sum_{y\in S(m,x)} p(m,y)}{p(m)}$, $\forall x,m$. We say $(p(M|Y), p(M), p(X|M))$ are* self-consistent solutions to **P2a**, **P2b**, **P2c** *if $\sum_{m\in S(x,y)} p(m|y) = p(x|y)$, $\forall x,y$ also holds. They are* self-consistent optimal solutions *if in addition the three components are optimal solutions to the subproblems **P2a**, **P2b**, **P2c**, respectively, where the constants specifying each subproblem instance are $p(X,Y)$ and the values of the other two components.*

A self-consistent set of solutions to these subproblems will correspond to a feasible solution to **P1** of the same cost (proof omitted). Unfortunately, $\mathcal{F}^3$ in general is *not* a convex function. We can prove the following, however.

Lemma 2. *$\mathcal{F}^3$ is strictly convex with respect to each of variable $p(m)$, $p(m|y)$, and $p(x|m)$* separately.

Theorem 2. *$\mathcal{F}^{M|Y}$ is convex.*

Algorithm 1 Iterative Algorithm for the Pareto Frontier (parameters $\beta \geq 0, \epsilon > 0$)

1: initialize $p(m|y)$ to arbitrary nonnegative values satisfying $\sum_{m\in S(x,y)} p(m|y) = p(x|y),\ \forall\ m,x,y$
2: **repeat**
3: $p_{\text{old}}(m|y) \leftarrow p(m|y)$ $\quad \forall m,y$
4: $p(m) \leftarrow \sum_y p(m|y)p(y)$ $\quad \forall m$
5: $p(x|m) \leftarrow \frac{\sum_{y\in S(m,x)} p(m|y)p(y)}{p(m)}$ $\quad \forall m,x$
6: $Z(x,y) \leftarrow \frac{1}{p(x|y)} \cdot \sum_{m\in S(x,y)} \left(p(m)\left(\frac{p(x|m)}{p(x)}\right)^{\beta}\right)$ $\forall x,y$
7: $p(m|y) \leftarrow \frac{1}{Z(x_{m,y},y)} \cdot \left(p(m)\left(\frac{p(x_{m,y}|m)}{p(x_{m,y})}\right)^{\beta}\right)$ $\quad \forall m,y$
8: **until** $JS_{\frac{1}{2},\frac{1}{2}}[p(m|y)\|p_{\text{old}}(m|y)] < \epsilon,\ \forall\ y$
9: **return** $p(m|x,y) = \begin{cases} p(m|y) & \text{if } x = x_{m,y} \\ 0 & \text{otherwise} \end{cases}$ $\forall\ m,x,y$

We find the explicit expressions for the iterative algorithm by solving **P2a**, **P2b**, **P2c** by the Lagrangian method (derivations omitted). These self-consistent equations become the assignments in lines 7, 4, 5 of Algorithm 1, respectively. For the exit condition we use the Jensen-Shannon (JS) divergence, defined as $JS_\Pi[p1,p2] := \pi_1 D_{KL}[p1\|\bar{p}] + \pi_2 D_{KL}[p2\|\bar{p}]$, where $\Pi = \{\pi_1,\pi_2\}, 0 < \pi_1,\pi_2 < 1, \pi_1+\pi_2 = 1$ and $\bar{p} = \pi_1 p_i + \pi_2 p_2$. In contrast to KL-divergence, JS-divergence is a bounded divergence and is less sensitive to low probability values.

We state the correctness of this algorithm as a theorem.

Theorem 3. *In the limit as parameter $\epsilon \to 0$, Algorithm 1 converges to a solution $(p(M|Y), p(M), p(X|M))$ with the following properties:*

a) it is a stationary point for $\mathcal{F}^3$;
*b) its three components are self-consistent solutions to **P2a**, **P2b**, **P2c**;*
*c) its component $p(M|Y)$ is an optimal solution to $\mathcal{F}^{M|Y}$, and thus to **P1**.*

Proof. a) When line 7 executes, it ensures that the constraints of **P2a** hold, implying (via Lemma 2) that the resulting $p(M|Y)$ is an optimal solution to the instance of **P2a** specified by the then-current solutions of **P2b** and **P2c**.

Line 4 is simplified from an equation derived in the full version of the paper, using the following property obtained from executing line 7 (or, in the first iteration, from the initial values):

$$\sum_{m} p(m|y) = 1, \quad \forall\ y \quad (2)$$

Therefore line 4 produces an optimal solution $p(M)$ to the instance of **P2b** specified by the then-current solutions of **P2c** and **P2a**.

Finally, line 5 is simplified from an equation derived in the full version of the paper, using the following property obtained from executing line 4:

$$p(m) = \sum_{y} p(m|y)p(y), \quad \forall\ m \quad (3)$$

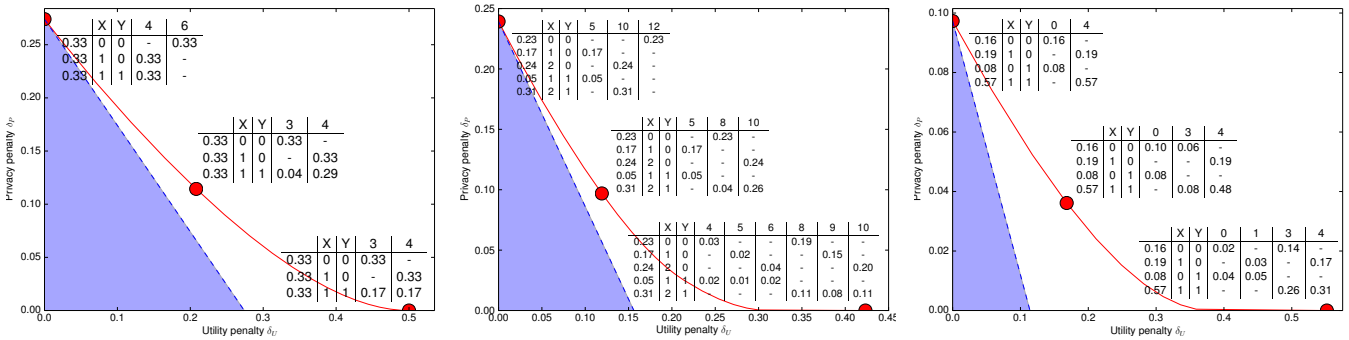


Fig. 2: Pareto frontiers for three sample problem instances. Shaded region indicates region known to be infeasible due to the lower bound of [1]. For each, the corresponding message distributions are shown for the two extreme points and one intermediate point. Notice that the lower bound is always achievable in the perfect utility case (upper left), simply by setting $M = X$ but is generally *not* achievable in the perfect utility case (lower right).

Thus line 5 produces an optimal solution $p(X|M)$ to the instance of **P2c** specified by the then-current solutions of **P2a** and **P2b**.

Therefore each execution of lines 7, 4, or 5 can only reduce the cost of their common objective function. From this and the fact that $\mathcal{F}^3$ is bounded from below (see full version of the paper), it follows that the algorithm must converge. This implies that the resulting solution $(p(M|Y), p(M), p(X|M))$ is a stationary point for $\mathcal{F}^3$.

b) Because lines 4 and 5 are recomputing $(p(M), p(X|M))$ based on $p(M|Y)$, the three components are self-consistent solutions.

c) Therefore $p(M|Y)$ *as a solution to* $\mathcal{F}^{M|Y}$ has the same value as the solution $(p(M|Y), p(M), p(X|M))$ for $\mathcal{F}^3$, and hence (because $\mathcal{F}^3$ is a *relaxation* of **P1**) $p(M|Y)$ is also a stationary point for $\mathcal{F}^{M|Y}$. Because $\mathcal{F}^{M|Y}$ is convex, it then follows that $p(M|Y)$ is an optimal solution to $\mathcal{F}^{M|Y}$, and thus to **P1** as well. □

V. Discussion

In this work, we took the first steps towards creating a theoretical framework for understanding the interplay between privacy and utility that arise in protecting a blacklist of unwanted inferences and simultaneously encouraging a whitelist of useful inferences. There are a number of other aspects of the problem that we intend to investigate in future work.

Computation complexity. Our algorithms in this paper for the privacy/utility problem, like those for IB and for the privacy filter/funnel problems, are non-polynomial time. These problems seem likely to be NP-hard, but to the best of our knowledge this question remains open. In [11] such problems are noted to be NP-hard "in general," citing the closely related rate-distortion problem, which indeed is known to be NP-hard [12]. There the goal is to minimize the average distortion $E[D(x, m)]$, based on a given distortion function $D(\cdot, \cdot)$, subject to a rate constraint. Also known to be hard [13] is a still more similar quantizing problem whose goal is to maximize the mutual information $I(X; M)$, again subject to a bound on $|\mathcal{M}|$. Although there are certain *asymptotic* equivalences between constraints on $|\mathcal{M}|$ and mutual information, this does not render hardness for these mutual information-based problems trivial.

Scalability. As already noted, the exponential size (in $|\mathcal{X}|$) of the message space poses a scalability challenge. One possible direction to explore is to coarsen the message space, by merging messages together.

Acknowledgement. This research is based upon work supported in part by the U.S. Army Research Laboratory and the U.K. Ministry of defense under Agreement Number W911NF-16-3-0001; by the NSF under awards INSPIRE-1547205, and by a CUNY Junior Faculty Research Award. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the NSF, the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

[1] S. Chakraborty, N. Bitouze, M. B. Srivastava, and L. Dolecek, "Protecting data against unwanted inferences," in *ITW*, 2013.

[2] A. A. Ali, S. M. Hossain, K. Hovsepian, M. M. Rahman, K. Plarre, and S. Kumar, "mpuff: Automated detection of cigarette smoking puffs from respiration measurements," ser. IPSN, 2012, pp. 269–280.

[3] P. Marquardt, A. Verma, H. Carter, and P. Traynor, "(sp)iphone: Decoding vibrations from nearby keyboards using mobile phone accelerometers," ser. CCS, 2011, pp. 551–562.

[4] E. Miluzzo, A. Varshavsky, S. Balakrishnan, and R. R. Choudhury, "Tapprints: Your finger taps have fingerprints," ser. MOBISYS, 2012, pp. 323–336.

[5] M. Yan, B. Dan, and N. Gabi, "Gyrophone: Recognizing speech from gyroscope signals," in *USENIX Security*, 2014, pp. 1053–1067.

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

[7] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," ser. Allerton Conference on Communication, Control and Computing, Allerton, IL, 1999, pp. 368–377.

[8] F. du Pin Calmon and N. Fawaz, "Privacy against statistical inference," *CoRR*, vol. abs/1210.2123, 2012.

[9] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Medard, "From the information bottleneck to the privacy funnel," *arXiv preprint arXiv:1402.1774*, 2014.

[10] L. Wasserman and S. Zhou, "A statistical framework for differential privacy," *JASA*, vol. 105, no. 489, pp. 375–389, 2009.

[11] N. Slonim, "The information bottleneck: Theory and applications," Ph.D. dissertation, Hebrew University of Jerusalem, 2002.

[12] M. R. Garey, D. S. Johnson, and H. S. Witsenhausen, "The complexity of the generalized lloyd - max problem," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 255–256, 1982.

[13] B. Mumey and T. Gedeon, "Optimal mutual information quantization is np-complete," in *Neural Information Coding (NIC) workshop poster, Snowbird UT*, 2003.