Information-Theoretic Private Interactive Mechanism

by

Bahman Moraffah

A Qualification Examination Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

ABSTRACT

An information-theoretic mechanism for privacy-guaranteed interactions is introduced between two memoryless correlated sources where each source is characterized by a pair of public and private variables. The interactions are modeled as a collection of $K/2$ pairs of random mappings, one pair for each of the K rounds of interactions. The $K/2$ random mapping pairs are chosen jointly to minimize the information leakage (privacy measure) over K rounds of the private variable of each source at the other source while ensuring that a desired measure of utility (distortion) of the revealed public variable is satisfied. Arguing that an average case information-theoretic privacy metric can be appropriate for streaming data settings, this paper shows that in general, interaction reduces privacy leakage by drawing some parallels between this problem and the classic interactive source coding problem. Specifically, for the log-loss distortion metric it is shown that the resulting interaction problem is an analog of an interactive information bottleneck problem for which a one-shot interactive mechanism is, in general, not optimal. For the resulting problem with a non-convex constraint space, an algorithm that extends the one-way agglomerative information bottleneck algorithm to the interactive setting is introduced.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

Consider an electric power system in which systems operators that manage specific sub-areas of the network share measurements with each other to obtain precise estimates of the underlying system state, i.e., complex voltages. Despite the need for such sharing and the value of high fidelity state estimates, such sharing is often limited due to privacy considerations; in the process of sharing measurements the operators do not wish to leak information about a subset of their internal states. However, since the measurements need to be shared, and often multiple times due to the iterative nature of power systems state estimation, it is crucial to understand: (a) the effect of applying privacy-preserving mechanisms on both the utility of estimation and leakage of the private data; and (b) the effect of multiple rounds of interaction and sharing on the net leakage.

Privacy in such a distributed "competitive" context is different from the traditional statistical database privacy setting in which data is published to ensure statistical value while ensuring that the privacy of any individual in the database is not comprised. In this database context, differential privacy with guarantees on the worst-case privacy leakage has emerged as a strong formalism [1]. However, in many data sharing settings, such as the above-mentioned electric power system example as well as other streaming data settings (e.g., sensors networks, IoT, even electronic medical records, etc), the data stream as a whole has private and public features that need to be hidden and revealed, respectively. In such settings, a statistical approach using mutual information as a privacy metric is more meaningful in quantifying information leakage and providing guarantees.

To this end, we consider a two-way interactive data sharing setting with two agents. Each agent generates an n-length independent and identically distributed (i.i.d.) sequence of public and private data; data at the two agents are assumed to be correlated as is generally the case in such distributed settings. Each agent wishes to share a function of its public data with the other agent to satisfy a desired measure of utility (e.g., via a distortion function) while ensuring that a mutual information based leakage of its private data is constrained over K rounds of communications. This problem model lends itself naturally to a rate-distortion based formulation. However, the problem at hand does not involve a rate constraint, and therefore does not require encoders and decoders; on the other hand, hiding private features from correlated public features in an interactive setting require a collection of random mappings, one for each round.

Formally, an information-theoretic privacy mechanism is a randomizing function that maps the public data from a data source to an output (*revealed/released* data); any such mapping will achieve a certain utility, quantified via a desired distortion function, and leakage of private data quantified via average mutual information. In the interactive setting, we allow for a total of K rounds of data sharing ($K/2$ rounds per agent) and introduce a private interactive mechanism as a collection of K random mappings. From both a theoretical and an application viewpoint, it is of much interest to understand whether interaction reduces privacy leakage or if a single round of data sharing suffices for a fixed privacy budget (leakage constraint). Furthermore, in contrast to the traditional interactive source coding setup, here the leakage and distortion constraints are on different aspects of the source, namely, the private and public features, respectively. Thus, it is unclear *a priori* if multiple rounds of interaction reduce leakage or worsen it.

Our Contributions: In this paper, we consider discrete memoryless correlated sources at the two agents and determine the set of all possible leakage-distortion tuples achievable at both agents over K rounds of interaction (Section 2). In addition to providing examples of sources for which interaction reduces leakage (Section 2.2), we focus on a specific class of distortion functions, namely, log-loss distortion (Section 3). Our motivation for this model stems from the fact that the soft decoding characteristic of many iterative systems is well captured by log-loss distortion. We show that the resulting problem is a dual of an interactive information bottleneck problem, and analogously, involves optimization over a non-convex probability space; to this end, we introduce a generalization of the agglomerative information bottleneck algorithm for the two-agent interactive case (Section 3.1) and illustrate the value of interaction in reducing leakage. Finally for Gaussian sources with both mean-squared and log-loss distortion, we prove the optimality of one-shot data sharing (Sections 3 and 3.1).

Related Work: An information-theoretic formulation of the utility-privacy tradeoff problem was introduced in [2] for the one-shot data publishing setting and has also been studied in [3, 4]. For the interactive setting, [5] determines the largest achievable utility-privacy tradeoff region for a two-agent system with a class of correlated Gaussian sources and mean-squared distortion functions at both agents. In contrast, the focus in this paper is on general source distributions and distortions.

For a one-way non-interactive setting, in [6] Makhdoumi *et al.* introduce an algorithm based on the agglomerative information bottleneck algorithm to compute the risk-distortion tradeoff for logarithmic loss based privacy and distortion functions. More recently, in [7] Vera *et al.* study the rate-relevance region for an interactive two-agent information bottleneck problem. In contrast to the information bottleneck problem in which the goal is to minimize the compression rate of one feature (considered public in our model) while ensuring the output guarantees a lower bound on the (mutual) information of a correlated feature (considered private in our model), under log-loss distortion, the problem we solve is a dual problem of minimizing information leakage of the hidden feature while lower bounding the (mutual) information of the public feature (log-loss distortion constraint); for this problem, we develop an algorithmic solution and highlight the advantages of multiple rounds of data sharing to reduce leakage.

It is worth noting that the problem at hand also falls under the purview of secure multiparty computation (SMC); in this context, recently, in [8] Kairouz *et al.* prove the optimality of one-shot interactions in a SMC setting using differentially private data sharing. While SMC is a compelling formal framework for secure distributed data sharing, we argue for alternate approaches due to both the complexity of practical SMC implementations, if and when possible, as well as our focus on problems wherein there is a need for data sharing without any central agent in a repeated fashion.

Chapter 2

SYSTEM MODEL AND INTERACTIVE MECHANISM

We consider two-way interactive model as shown in Fig. 2.1, where agents A and B generate n-length i.i.d. sequences (X_1^n, Y_1^n) and (X_2^n, Y_2^n), respectively, with $(X_{1i}, Y_{1i}, X_{2i}, Y_{2i}) \sim P_{X_1,Y_1,X_2,Y_2}$, for all $i = 1, 2, ..., n$. The public data at both agents are denoted by $X_{(\cdot)}^n$ and the correlated private data by $Y_{(\cdot)}^n$. Furthermore, we assume that the private data is hidden and can only be leaked through the public data. We consider a K-round interactive protocol in which, without loss of generality, we assume that agent A initiates the interaction and K is even. A K-interactive privacy mechanism is given by $(n, K, \{P_{1i}\}_{i=1}^{K/2}, \{P_{2i}\}_{i=1}^{K/2}, D_1, D_2, L_1, L_2)$ as a collection of K probabilistic mappings such that agent A shares data in the odd rounds beginning with round 1 and agent B shares in the even rounds. A privacy mechanism P_{1i} for agent A used in the $(2i-1)$-th round, $i \in \{1, 2, \ldots, \frac{K}{2}\}$, is a mapping from its public data sequence and all prior sequences revealed from agent B. Thus, in round 1, $P_{11} : \mathcal{X}_1^n \to \mathcal{U}_1^n$, where $\mathcal{U}_1^n$ is the revealed set when a sequence U_1^n is shared via P_{11}. For the odd rounds $i = 3, \ldots, K-1$, the mechanism used by agent A is

$$P_{1,\frac{i+1}{2}} : (\mathcal{X}_1^n, \mathcal{U}_1^n, \mathcal{U}_2^n, \ldots, \mathcal{U}_{i-1}^n) \to \mathcal{U}_i^n. \tag{2.1}$$

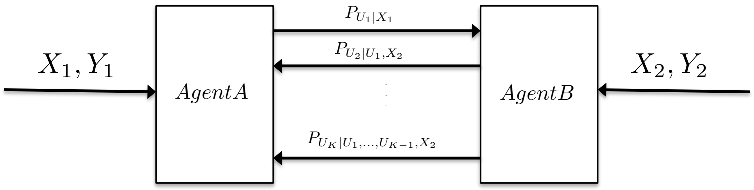

Figure 2.1: K-interactive Privacy Model.

Similarly, agent B in even rounds i, $i \in \{2, 4, \dots, K\}$ uses its public data and the prior data sequences revealed from agent A and maps them via a privacy mechanism

$$P_{2,\frac{i}{2}} : (\mathcal{X}_2^n, \mathcal{U}_1^n, \dots, \mathcal{U}_{i-1}^n) \to \mathcal{U}_i^n. \tag{2.2}$$

At the end of K rounds, agents A and B reconstruct sequences $\hat{X}_2^n$ and $\hat{X}_1^n$, respectively, where $\hat{X}_1^n = g_2(X_2^n, U_1^n, \dots, U_K^n)$ and $\hat{X}_2^n = g_1(X_1^n, U_1^n, \dots, U_K^n)$, and g_1 and g_2 are appropriately chosen functions. The set of mechanism pair $\{P_{1j}, P_{2j}\}_{j=1}^{\frac{K}{2}}$ is chosen to satisfy

$$\frac{1}{n}\sum_{i=1}^{\infty} E(d_1(X_{1i}, \hat{X}_{1i})) \leq D_1 + \epsilon \tag{2.3a}$$

$$\frac{1}{n}\sum_{i=1}^{\infty} E(d_2(X_{2i}, \hat{X}_{2i})) \leq D_2 + \epsilon \tag{2.3b}$$

$$\frac{1}{n} I(Y_1^n; U_1^n, \dots, U_K^n, X_2^n) \leq L_1 + \epsilon \tag{2.3c}$$

$$\frac{1}{n} I(Y_2^n; U_1^n, \dots, U_K^n, X_1^n) \leq L_2 + \epsilon \tag{2.3d}$$

where $d_1(\cdot,\cdot)$ and $d_2(\cdot,\cdot)$ are the given distortion measures.

From the problem definition, it follows that $Y_1 \leftrightarrow X_1 \leftrightarrow \hat{X}_1$, $Y_2 \leftrightarrow X_2 \leftrightarrow \hat{X}_2$ form Markov chains. The utility-privacy tradeoff region is the set of all (L_1, D_1, L_2, D_2) tuples for which a privacy mechanism exists.

Theorem 1. *For target distortion pair (D_1, D_2), and for a K-round interactive privacy mechanism the utility-privacy tradeoff region is given as:*

$$\begin{aligned} \{(L_1, L_2, D_1, D_2) : L_1 &\geq I(Y_1; U_1, \dots, U_K, X_2), \\ L_2 &\geq I(Y_2; U_1, \dots, U_K, X_1), \\ E(d_1(X_1, \hat{X}_1)) &\leq D_1, \\ E(d_2(X_2, \hat{X}_2)) &\leq D_2\} \end{aligned} \tag{2.4}$$

such that for all k, the following Markov chains hold:

$$Y_1 \leftrightarrow (U_1, \ldots, U_{2k-1}, X_2) \leftrightarrow U_{2k} \tag{2.5}$$

$$Y_2 \leftrightarrow (U_1, \ldots, U_{2k-2}, X_1) \leftrightarrow U_{2k-1} \tag{2.6}$$

with $|\mathcal{U}_l| \leq |\mathcal{X}_{i_l}|.(\prod_{j=1}^{l-1} |\mathcal{U}_j|) + 1$ *where* $i_l = 1$ *if* l *is odd and* $i_l = 2$ *if* l *is even.*

Proof. The proof details are in Appendix A. We briefly review the steps. The proof involves two steps: the achievability part uses the method of types and typicality arguments in bounding the achievable leakages; the converse on the other hand, considers a mechanism that achieves (2.3a)-(2.3d) and exploits the i.i.d. nature of correlated sources to obtain single letter bounds. □

Corollary 1. *For the special case,* $Y_i = X_i$*,* $i = 1, 2$*, i.e., the public and private data are the same, the leakage-distortion region in Theorem 1 is the same as the rate-distortion region for the interactive source coding problem in[9].*

Remark 1. *Note that a one-shot setting is one in which both agents share data independently and simultaneously with each other only once.*

Without loss of generality we assume we initiate interaction from agent A such that the last round of interaction is from agent B to agent A. We define a compact subset of a finite Euclidean space as

$$\begin{aligned}\mathcal{P}_K^A := \{& P_{U^K|X_1,Y_1,X_2,Y_2} : P_{U^K|X_1,Y_1,X_2,Y_2} = P_{U_1|X_1} P_{U_2|U_1,X_2} \ldots, P_{U_K|U^{K-1},X_2}, \\ & E(d_1(X_1, \hat{X}_1)) \leq D_1, E(d_1(X_2, \hat{X}_2)) \leq D_2\} \end{aligned} \tag{2.7}$$

In addition to the tradeoff region, one can also focus on the net leakage over K rounds. From Theorem 1, the sum leakage-distortion function initiates from agent A over K rounds is

$$L^A_{sum,K}(D_1, D_2) = \min_{P_{U^K|X_1,Y_1,X_2,Y_2} \in \mathcal{P}^A_K} \{I(Y_1; U_1, \ldots, U_K, X_2) + I(Y_2; U_1, \ldots, U_K, X_1)\}. \tag{2.8}$$

For the region given by Theorem 1, one can define a sum leakage over any k rounds, $k = 1, 2, \ldots, K$ with target distortions D_1 and D_2. Depending on which agent initiates the interactions (assuming agent A), we have

$$\begin{aligned} L^A_{sum,k}(D_1, D_2) = &\sum_{\substack{i,j=1 \\ i\neq j}}^{2} I(Y_i, X_j) \\ &+ \min_{P_{U^K|X_1,Y_1,X_2,Y_2} \in \mathcal{P}^A_K} \left(\sum_{i=1}^{k} I(Y_1; U_i | X_2, U^{i-1}) + \sum_{i=1}^{k} I(Y_2; U_i | X_1, U^{i-1}) \right). \end{aligned} \tag{2.9}$$

One can similarly define $L^B_{sum,k}$ for sum leakage over k rounds originating from agent B.

Lemma 1. *For all k (I) $L^A_{sum,(k-1)} \geq L^A_{sum,k}$. Similarly, $L^B_{sum,(k-1)} \geq L^B_{sum,k}$. (II) $L^B_{sum,(k-1)} \geq L^A_{sum,k}$. Similarly, $L^A_{sum,(k-1)} \geq L^B_{sum,k}$.*

Proof. (I) For all k, $L^A_{sum,(k-1)} \geq L^A_{sum,k}$, since any $(k-1)$-round interactive mechanism with initial at agent A can be considered as special case of k-round interactive mechanism with initial at agent A and $P_{U_k|\cdot,\cdot,\cdot,\cdot} = 0$. (II) For all k, $L^B_{sum,(k-1)} \geq L^A_{sum,k}$, since any $(k-1)$-round interactive mechanism initial at A can be considered as special case of k-round interactive mechanism with initial at B with $P_{U_1|X_1} = 0$. □

Definition 1. $L_{sum,\infty} := \lim_{k\to\infty} L^A_{sum,k} = \lim_{k\to\infty} L^B_{sum,k}$.

From (II) Lemma 1, $\lim_{k\to\infty} L^A_{sum,k} = \lim_{k\to\infty} L^B_{sum,k}$. From (I) Lemma 1, $L^A_{sum,k}$, $L^B_{sum,k}$ are non-increasing in k and bounded from below, so limit exists. Thus, $L_{sum,\infty}$ is well-defined.

Remark 2. *Note that Theorem 1 holds even if agent B initiates the interaction; however now U_1 will be the output of agent B in round 1 and U_2 will be the output of agent A in round 2, and so on, such that the Markov conditions are appropriate in Theorem 1.*

2.1 When Does Interaction help?

A natural question to ask that follows the characterization of the k-round leakage-distortion region is whether interaction actually reduces leakage relative to a one-round setting. In this section, we introduce a test for checking when multiple rounds of interaction help by first identifying the relationship between $L_{sum,\infty}$ and $L^A_{sum,K}$ and then using it to determine the conditions under which interaction reduce leakage.

Our approach is modeled along the lines of the method in [10] by Ma *et al.* in which an interactive source coding problem is considered. However, since our source models include a pair of public and private variables, we need to extend the methods in [10] to the problem setting at hand. Specifically, we characterize $L_{sum,\infty}$ and compare it with $L^A_{sum,k}$ for any given k to determine the value of interaction.

The characterization of $L^A_{sum,k}$ in (2.8) does not give us any bounds on the rate of convergence to $L_{sum,\infty}$ for a given distribution P_{X_1,Y_1,X_2,Y_2}. Note that for each finite k, as k increases, the dimension of optimization problem in (2.8) explodes. In this section, we tackle this problem differently. Instead of determining the characterization of $L_{sum,\infty}$ for a fixed joint distribution of P_{X_1,Y_1,X_2,Y_2} and taking a limit as $t \to \infty$, we characterize $L_{sum,\infty}$ for a family of distributions.

Without loss of generality, let agent A initiate a K-round interaction with agent B such that the last round of interaction is from agent B to agent A. The goal is to characterize the family of source distributions for which interaction helps. To this end, we define a "leakage reduction" function $\eta^A_K(P_{X_1,Y_1,X_2,Y_2}, D_1, D_2)$.

Definition 2. *The leakage reduction function for a K-round interactive mechanism initiated at agent A is defined as*

$$\begin{aligned}\eta_K^A(P_{X_1,Y_1,X_2,Y_2},D_1,D_2) &:= H(Y_1)+H(Y_2)-L^A_{sum,K}(D_1,D_2)\\ &= \max_{P(u^K|x_1,y_1,x_2,y_2)\in\mathcal{P}_K^A}[H(Y_1|U^K,X_2)+H(Y_2|U^K,X_1)] \qquad (2.10)\end{aligned}$$

Note that $\eta_K^A(P_{X_1,Y_1,X_2,Y_2},D_1,D_2)$ depends on P_{X_1,Y_1,X_2,Y_2} only through $P_{X_1,Y_1|X_2}$ and $P_{X_1,Y_1|X_2}$. Evaluating η_K^A is equivalent to evaluating $L^A_{sum,K}$. Definition 2 enables us to characterize the properties of $\eta_\infty = \lim_{K\to\infty}\eta_K^A$ which then gives us $L^A_{sum,\infty} = H(Y_1)+H(Y_2)-\eta_\infty$. The goal is to determine source distributions for which $\eta_\infty \le \eta_0$ where η_0 is the leakage reduction the absence of interaction. When $K=0$, we have $L^A_{sum,0} = L^B_{sum,0} = L_{sum,0} = I(Y_1;X_2)+I(Y_2;X_1)$ and consequently, $\eta_0 = H(Y_1|X_2)+H(Y_2|X_1)$.

Generally, η_K^A and $L^A_{sum,K}$ are functionals of P_{X_1,Y_1,X_2,Y_2}, D_1, and, D_2. For a given source, since it is generally not possible to precisely determine the rate of convergence of $L^A_{sum,k}$ to $L_{sum,\infty}$, we focus, as in [10], on determining the set of source distributions for which $L_{sum,\infty}$ is strictly decreasing. This leads us to define the set of structured neighborhoods of P_{X_1,Y_1,X_2,Y_2} which is the collection of all joint distribution P'_{X_1,Y_1,X_2,Y_2} which have the same marginal $P_{X_2,Y_2|X_1}$ as follows.

Definition 3. *The marginal perturbation set $\mathcal{P}_{X_2,Y_2|X_1}$ for a given joint distribution P_{X_1,Y_1,X_2,Y_2} is defined as*

$$\mathcal{P}_{X_2,Y_2|X_1}(P_{X_1,Y_1,X_2,Y_2}) = \{P'_{X_1,Y_1,X_2,Y_2} : P'_{X_1,Y_1,X_2,Y_2} << P_{X_1,Y_1,X_2,Y_2}, P'_{X_2,Y_2|X_1} = P_{X_2,Y_2|X_1}\} \qquad (2.11)$$

where " $<<$ " is majorizing operator. This set is an ordered set with respect to majorization. One can similarly define $\mathcal{P}_{X_1,Y_1|X_2}(P_{X_1,Y_1,X_2,Y_2})$.

Remark 3. *Note that $\mathcal{P}_{X_2,Y_2|X_1}(P_{X_1,Y_1,X_2,Y_2})$ and $\mathcal{P}_{X_1,Y_1|X_2}(P_{X_1,Y_1,X_2,Y_2})$ are nonempty* *sets as they contain P_{X_1,Y_1,X_2,Y_2}. Furthermore, for all P_{X_1,Y_1,X_2,Y_2}, $\mathcal{P}_{X_2,Y_2|X_1}(P_{X_1,Y_1,X_2,Y_2})$ and $\mathcal{P}_{X_1,Y_1|X_2}(P_{X_1,Y_1,X_2,Y_2})$ are convex sets of P_{X_1,Y_1,X_2,Y_2}.*

We now develop characterization of η_∞ for all $P_{X_1,Y_1,X_2,Y_2} \in \mathcal{P}_{X_1,Y_1,X_2,Y_2}$ defined as follows which is closed with respect to marginal perturbation.

Definition 4. *A family of joint distributions $\mathcal{P}_{X_1,Y_1,X_2,Y_2}$ is marginal-perturbation-closed if for all $P_{X_1,Y_1,X_2,Y_2} \in \mathcal{P}_{X_1,Y_1,X_2,Y_2}$, $\mathcal{P}_{X_2,Y_2|X_1} \cup \mathcal{P}_{X_1,Y_1|X_2} \subseteq \mathcal{P}_{X_1,Y_1,X_2,Y_2}$.*

To characterize η_∞, we define the following family of functionals.

Definition 5. *η_0-majorizing family of functionals $\mathcal{F}_D(\mathcal{P}_{X_1,Y_1,X_2,Y_2})$ is the set of all functionals $\eta : \mathcal{P}_{X_1,Y_1,X_2,Y_2} \times \mathcal{D}^2 \to \mathbb{R}$ satisfying*

1. *For all $P_{X_1,Y_1,X_2,Y_2} \in \mathcal{P}_{X_1,Y_1,X_2,Y_2}$ and $(D_1, D_2) \in \mathcal{D}^2$, $\eta(P_{X_1,Y_1,X_2,Y_2}, D_1, D_2) \geq \eta_0(P_{X_1,Y_1,X_2,Y_2}, D_1, D_2)$.*

2. *For all $P_{X_1,Y_1,X_2,Y_2} \in \mathcal{P}_{X_1,Y_1,X_2,Y_2}$, η is concave on $\mathcal{P}_{X_2,Y_2|X_1}$.*

3. *For all $P_{X_1,Y_1,X_2,Y_2} \in \mathcal{P}_{X_1,Y_1,X_2,Y_2}$, η is concave on $\mathcal{P}_{X_1,Y_1|X_2}$.*

To characterize the properties of η_∞ we need to establish the relationship between $(k-1)$-round interactive mechanism and k-round interactive mechanism. Intuitively speaking, to construct a k-round interactive mechanism, we first pick U_1, then for each realization of $U_1 = u_1$ constructing the remaining by considering $(k-1)$-round initiated at agent B but with different data distribution $P_{X_1,Y_1,Y_1,Y_2|U_1=u_1} \in \mathcal{P}_{X_2,Y_2|X_1}(P_{X_1,Y_1,X_2,Y_2})$. Distortion vector $(D_1', D_2)_{u_1}$ for each realization $U_1 = u_1$ in $(k-1)$-round interactive subproblem could be different from the original distortion vector (D_1, D_2). The only condition needs to be satisfied is $\sum_{u_1} (D_1', D_2)_{u_1} P_{U_1}(u_1) = (D_1, D_2)$. The following lemma will be used in determining the η_∞.

Lemma 2. *1. For all $k \in \mathbb{Z}^+$ and $P_{X_1,Y_1,X_2,Y_2} \in \mathcal{P}_{X_1,Y_1,X_2,Y_2}$ we have*

$$\eta_k^A(P_{X_1,Y_1,X_2,Y_2}, D_1, D_2) = \max_{P(U_1|X_1)} \left\{ \max_{\substack{\forall u_1 \in \mathcal{U}_1, (D_1', D_2)_{u_1} \in \mathcal{D}^2 \\ (D_1', D_2)_{u_1} : E((D_1', D_2)_{u_1}) \leq (D_1, D_2)}} \{ \sum_{u_1 \in \mathcal{U}_1} g(u_1) \} \right\}. \tag{2.12}$$

where $g(u_1) = P_{U_1}(u_1)\eta_{k-1}^B(P(X_1, Y_1, X_2, Y_2|u_1), (D_1', D_2)_{u_1})$.

2. For all $k \in \mathbb{Z}^+$ and all $(q_{X_1,Y_1,X_2,Y_2}, D_1, D_2) \in \mathcal{P}_{X_1,Y_1,X_2,Y_2} \times \mathcal{D}^2$, η_k^A is concave on $\mathcal{P}_{X_2,Y_2|X_1} \times \mathcal{D}^2$.

3. For all $k \in \mathbb{Z}^+$ and all $(q_{X_1,Y_1,X_2,Y_2}, D_1, D_2) \in \mathcal{P}_{X_1,Y_1,X_2,Y_2} \times \mathcal{D}^2$, if $\eta : \mathcal{P}_{X_1,Y_1,X_2,Y_2} \times \mathcal{D}^2 \to \mathbb{R}$ is concave on $\mathcal{P}_{X_2,Y_2|X_1} \times \mathcal{D}^2$ and if for all $(P_{X_1,Y_1,X_2,Y_2}, D_1, D_2) \in \mathcal{P}_{X_2,Y_2|X_1}(q_{X_1,Y_1,X_2,Y_2}) \times \mathcal{D}^2$,$\eta_{k-1}^B(P_{X_1,Y_1,X_2,Y_2}, D_1, D_2) \leq \eta(P_{X_1,Y_1,X_2,Y_2}, D_1, D_2)$, then for all $(P_{X_1,Y_1,X_2,Y_2}, D_1, D_2) \in \mathcal{P}_{X_2,Y_2|X_1}(q_{X_1,Y_1,X_2,Y_2}) \times \mathcal{D}^2$, $\eta_k^A(P_{X_1,Y_1,X_2,Y_2}, D_1, D_2) \leq \eta(P_{X_1,Y_1,X_2,Y_2}, D_1, D_2)$.

Proof. The proof is based on Lemma 1 in [10] and we provide a sketch below.

1. For all $k \in \mathbb{Z}^+$ and $P_{X_1,Y_1,X_2,Y_2} \in \mathcal{P}_{X_1,Y_1,X_2,Y_2}$

$$\begin{aligned}
&\eta_K^A(P_{X_1,Y_1,X_2,Y_2}, D_1, D_2) \\
&= \max_{P_{U^K|X_1,Y_1,X_2,Y_2} \in \mathcal{P}_K^A} [H(Y_1|U^K, X_2) + H(Y_2|U^K, X_1)] \\
&= \max_{P_{U_1|X_1}} \left\{ \max_{\substack{P_{U_2^K|X_1,Y_1,X_2,Y_2,U_1} : \\ P_{U_1|X_1} P_{U_2^K|X_1,Y_1,X_2,Y_2,U_1} \in \mathcal{P}_K^A}} [H(Y_1|U^K, X_2) + H(Y_2|U^K, X_1)] \right\} \\
&= \max_{P_{U_1|X_1}} \left\{ \max_{\substack{\forall u_1 \in \mathcal{U}_1, (D_1', D_2)_{u_1} \in \mathcal{D}^2 \\ (D_1', D_2)_{u_1} : E((D_1', D_2)_{u_1}) \leq (D_1, D_2)}} \left\{ \sum_{u_1} P_{U_1}(u_1) \left\{ \max_{\substack{P_{U_2^K|X_1,Y_1,X_2,Y_2,U_1} : \\ P_{U_1|X_1} P_{U_2^K|X_1,Y_1,X_2,Y_2,U_1} \in \mathcal{P}_K^A}} \right. \right. \right.
\end{aligned}$$

$$[H(Y_1|U_2^K, X_2, U_1 = u_1) + H(Y_2|U_2^K, X_1, U_1 = u_1)]\Big\}\Big\}\Big\}$$

$$= \max_{P(U_1|X_1)} \Bigg\{$$

$$\max_{\substack{\forall u_1 \in \mathcal{U}_1, (D_1', D_2)_{u_1} \in \mathcal{D}^2 \\ (D_1', D_2)_{u_1} : E((D_1', D_2)_{u_1}) \leq (D_1, D_2)}} \left\{ \sum_{u_1 \in \mathcal{U}_1} P(u_1) \eta_{k-1}^B (P(X_1, Y_1, X_2, Y_2 | u_1), (D_1', D_2)_{u_1}) \right\}\Bigg\} \tag{2.13}$$

2. For all $k \in \mathbb{Z}^+$ and all $(q_{X_1,Y_1,X_2,Y_2}, D_1, D_2) \in \mathcal{P}_{X_1,Y_1,X_2,Y_2} \times \mathcal{D}^2$, consider two arbitrary distributions $P^1_{X_1,Y_1,X_2,Y_2}, P^2_{X_1,Y_1,X_2,Y_2} \in \mathcal{P}_{X_2,Y_2|X_1}$ and distortion vectors $D^1 = (D^1_1, D^1_2), D^2 = (D^2_1, D^2_2) \in \mathcal{D}^2$. For every $\lambda \in (0,1)$, define $P^3_{X_1,Y_1,X_2,Y_2} = \lambda P^1_{X_1,Y_1,X_2,Y_2} + \bar{\lambda} P^2_{X_1,Y_1,X_2,Y_2}$ and $D^3 = \lambda D^1 + \bar{\lambda} D^2$. We show that $\eta_k^A(P^3_{X_1,Y_1,X_2,Y_2}, D^3) \geq \lambda \eta_k^A(P^1_{X_1,Y_1,X_2,Y_2}, D^1) + \bar{\lambda} \eta_k^A(P^2_{X_1,Y_1,X_2,Y_2}, D^2)$. Define an auxiliary random variable $V \in \mathcal{U}_1 \times \{1, 2\}$ such that $P_V(u_1, 2) = \bar{\lambda} P_{U_1^2}(u_1)$ and $P_V(u_1, 1) = \lambda P_{U_1^1}(u_1)$ where $P_{U_1^1}, P_{U_1^2}$ are distributions that maximize (2.10) for distributions $P^1_{X_1,Y_1,X_2,Y_2}, P^2_{X_1,Y_1,X_2,Y_2}$, respectively. According to part 1 of lemma, we have

$$\lambda \eta_k^A(P^1_{X_1,Y_1,X_2,Y_2}, D^1) + \bar{\lambda} \eta_k^A(P^2_{X_1,Y_1,X_2,Y_2}, D^2)$$
$$= \lambda \sum_{u_1} P_{U_1^1}(u_1) \eta_{k-1}^B (P^1_{X_1,Y_1,X_2,Y_2|u_1}, (D^1_1, D^1_2)_{u_1}))$$
$$+ \bar{\lambda} \sum_{u_1} P_{U_1^2}(u_1) \eta_{k-1}^B (P^2_{X_1,Y_1,X_2,Y_2|u_1}, (D^2_1, D^2_2)_{u_1}))$$
$$= \sum_{\substack{V, \\ i=1,2}} P_V(u_1, i) \eta_{k-1}^B (P^i_{X_1,Y_1,X_2,Y_2|u_1}, (D^i_1, D^i_2)_{u_1})) \leq \eta_k^A(P^3_{X_1,Y_1,X_2,Y_2}, D^3). \tag{2.14}$$

3. Part 3 follows directly from part 1.

□

Remark 4. *By reversing the roles of agent A and B in Lemma 2, one can prove the same lemma for agent B.*

Theorem 2. *$\eta_\infty(P_{X_1,Y_1,X_2,Y_2}, D_1, D_2) \in \mathcal{F}_\mathcal{D}(\mathcal{P}_{X_1,Y_1,X_2,Y_2})$ and η_∞ is the least element of the set $\mathcal{F}_\mathcal{D}(\mathcal{P}_{X_1,Y_1,X_2,Y_2})$.*

Proof. We show that η_∞ satisfies all three conditions in Definition 5 as follows:

1. Condition 1 in Definition 5 is satisfied since $L_{sum,\infty} \leq L_{sum,0}$, due to part (I) Lemma 1.

2. Condition 2 in Definition 5 is satisfied due to part 2 in Lemma 2.

3. Condition 3 in Definition 5 is satisfied due to Remark 4.

We now prove that η_∞ is the smallest element of $\mathcal{F}_\mathcal{D}(\mathcal{P}_{X_1,Y_1,X_2,Y_2})$: we need to show $\forall \eta \in \mathcal{F}_\mathcal{D}(\mathcal{P}_{X_1,Y_1,X_2,Y_2}) \times \mathcal{D}^2$, $\forall P_{X_1,Y_1,X_2,Y_2} \in \mathcal{P}_{X_1,Y_1,X_2,Y_2}$, and, $\forall k$, $\eta_k^A(P_{X_1,Y_1,X_2,Y_2}) \leq \eta(P_{X_1,Y_1,X_2,Y_2})$ and $\eta_k^B(P_{X_1,Y_1,X_2,Y_2}) \leq \eta(P_{X_1,Y_1,X_2,Y_2})$. By using induction on k, part 3 of Lemma 2, and, Remark 4 we can show that η_∞ is the least element of $\mathcal{F}_\mathcal{D}(\mathcal{P}_{X_1,Y_1,X_2,Y_2})$. □

Note that due to Lemma 2 and Remark 4, η_k^A always satisfies conditions 1 and 2 in Definition 5, but not necessarily condition 3. By Theorem 3, η_∞ satisfies all the conditions in Definition 5. Once η_K^A satisfies condition 3 in Definition 5, then interaction is not required, in other words, if all three conditions in Definition 5 are not satisfied, it is beneficial to increase the number of rounds. We now summarize this in the following theorem.

Theorem 3. *The following equivalent conditions establish when interaction does not help.*

1. *For all $P_{X_1,Y_1,X_2,Y_2} \in \mathcal{P}_{X_1,Y_1,X_2,Y_2}$ and $D = (D_1, D_2) \in \mathcal{D}^2$, $\eta_k^A(P_{X_1,Y_1,X_2,Y_2}, D) = \eta_\infty(P_{X_1,Y_1,X_2,Y_2}, D)$.*

2. *For all* $P_{X_1,Y_1,X_2,Y_2} \in \mathcal{P}_{X_1,Y_1,X_2,Y_2}$ *and* $D = (D_1, D_2) \in \mathcal{D}^2$, $\eta_k^A(P_{X_1,Y_1,X_2,Y_2}, D) = \eta_{k+1}^B(P_{X_1,Y_1,X_2,Y_2}, D)$.

3. *For all* $P_{X_1,Y_1,X_2,Y_2} \in \mathcal{P}_{X_1,Y_1,X_2,Y_2}$ *and* $D = (D_1, D_2) \in \mathcal{D}^2$, η_k^A *is concave on* $\mathcal{P}_{X_1,Y_1|X_2}(P_{X_1,Y_1,X_2,Y_2}) \times \mathcal{D}^2$.

Proof. Condition 1 implies condition 2 since $\eta_k^A \leq \eta_{k+1}^B \leq \eta_\infty$. This inequality holds due to (II) Lemma 1. Condition 2 implies condition 3 due to Remark 4. Condition 3 implies condition 1can be shown by using part 2 in Lemma 2 in addition to the fact that $\eta_k^A \geq \eta_0$, which leads to $\eta_k^A \in \mathcal{F}_{\mathcal{D}}(\mathcal{P}_{X_1,Y_1,X_2,Y_2})$. According to Theorem 2, since η_∞ is the least element of $\mathcal{F}_{\mathcal{D}}(\mathcal{P}_{X_1,Y_1,X_2,Y_2})$ we have $\eta_k^A \geq \eta_\infty$. Therefore, $\eta_k^A = \eta_\infty$. □

2.2 Interaction Reduces Leakage: Illustration

A natural question in the interactive setting is to understand whether multiple rounds can reduce leakage of the private variables while achieving the desired distortion. In general, it is unclear whether interaction would reduce leakage relative to a one-shot setting. We now present an example where interaction helps.

We observe that our example is similar to the one in [11] wherein Ma *et. al.* consider an interactive source coding problem for sources (X_1, X_2) at the two agents, i.e., without private data (Y_1, Y_2) and with constraints on coding rate in place of leakage. However, it is not clear the optimal mechanisms for the rate-distortion problem hold when minimizing leakage of (Y_1, Y_2). In fact, one needs to evaluate the optimal mechanism for the problem at hand in each round due to the presence of private side information at each agent and the leakage function being minimized; we detail these computations below.

We consider binary random variables X_1, X_2, Y_1, Y_2 such that (X_1, X_2) is modeled as doubly symmetric binary source with parameter p, i.e., $(X_1, X_2) \sim DSBS(p)$, with

$P_{X_1,X_2}(0,0) = P_{X_1,X_2}(1,1) = \frac{1-p}{2}$ and $P_{X_1,X_2}(1,0) = P_{X_1,X_2}(0,1) = \frac{p}{2}$. Furthermore, (X_1, Y_1) and (X_2, Y_2) are correlated as follows: $Y_1 = X_1 \oplus Z_1$ and $Y_2 = X_2 \oplus Z_2$ where $Z_i \sim Ber(p)$ for $i = 1, 2$, and Z_1 and Z_2 are independent of X_1 and X_2, respectively. We let $d_A = 0$ and consider an erasure distortion measure $d_B(\cdot, \cdot)$ as:

$$d_B(x_1, \hat{x}_1) = \begin{cases} 0, & \text{if } \hat{x}_1 = x_1 \\ 1, & \text{if } \hat{x}_1 = e \\ \infty, & \text{if } \hat{x}_1 = 1 - x_1. \end{cases} \tag{2.15}$$

One-round sum leakage $L^A_{sum,1}$: We first compute sum leakage $L^A_{sum,1}$ for a one round interaction starting from agent A. Note that in this case even though B does not share data, by definition, the sum leakage $L^A_{sum,1}$ includes the leakage of Y_2 at A. In the Appendix, we prove that the leakage-distortion function is

$$L^A_{sum,1}(0, D_2) = 2 - [(1 - D_2)H(p) + (1 + D_2)H(2p(1 - p))]. \tag{2.16}$$

For the classical source coding problem with the same distribution defined above for (X_1, X_2) and functional $d_B(\cdot, \cdot)$ in (2.15), the optimal $P_{U_1|X_1}$ minimizing the Wyner-Ziv rate-distortion function $I(X_1; U_1|X_2)$ is well known[12]. However, it is not clear *a priori* that the same transition probability distribution will also minimize the leakage $I(Y_1; U_1|X_2)$ in the presence of private features at both agents. In the Appendix, we prove that $I(Y_1; U_1, X_2)$ is indeed minimized by the same distribution that minimizes $I(X_1; U_1|X_2)$ and achieves the Wyner-Ziv rate-distortion function (without Y_i^n for $i = 1, 2$). This is also a result of independent interest.

Two-round sum leakage $L^B_{sum,2}$: We now compute the sum leakage $L^B_{sum,2}$ for a two-round interaction starting from agent B in round 1 and returning from A to B in round 2. Let U_1^n denote the output of the mapping in round 1 from B to A and U_2^n denotes the output of mapping in round 2 from A to B. We will explicitly construct a mechanism pair $(P_{U_1|X_2}, P_{U_2|X_1,U_1})$ and $\hat{X}_1$ which leads to an admissible tuple (L_1, L_2, D). Let $P_{U_1|X_2}$ be binary symmetric channel with crossover probability α, i.e., $P(U_1|X_2) = BSC(\alpha)$. We choose the conditional pmf $P_{U_2|X_1,U_1}(u_2|x_1,u_1)$ as given in Table 2.1 and let $\hat{X}_1 = U_2$.

Table 2.1: Conditional Distribution $P_{U_2|X_1,U_1}$

$P_{U_2\|X_1,U_1}$	$u_2 = 0$	$u_2 = e$	$u_2 = 1$
$x_1 = 0, u_1 = 0$	$1-\beta$	β	0
$x_1 = 1, u_1 = 0$	0	1	0
$x_1 = 0, u_1 = 1$	0	1	0
$x_1 = 1, u_1 = 1$	0	β	$1-\beta$

For a given value for the DSBS parameter, p, there are several values of (α, β) pair such that $L^B_{sum,2} \leq L^A_{sum,1}$. For example, for $p = 0.03$, $\alpha = 0.35$, and $\beta = 0.55$, $L^B_{sum,2}(0, D_2)$ is

$$I(Y_2; U_1, X_1) + I(Y_1; U_2|U_1, X_2) = 1.1876 \tag{2.17}$$

and the corresponding distortion is $D_2 = E(d(X_1, \hat{X}_1)) = 0.8116$. By computing $L^A_{sum,1}$ and comparing it with (14) for the same distortion, we have $L^A_{sum,1}(0, 0.8116) = 1.3832$. Thus, interaction reduces leakage.

In [11], using the same $P_{U_1|X_2}$, $P_{U_2|X_1,U_1}(u_2|x_1,u_1)$ and $\hat{X}_1$ as described above, Ma *et. al.* show that interaction reduces the sum-rate over two rounds relative to one round for specific values of p, α, and β. However, as discussed earlier, it wasn't clear whether the same parameters in [11] also reduce leakage of correlated hidden variables (Y_1, Y_2) in our problem. We have verified that for different value of α and β including those in [11], the two-round sum leakage is smaller than the one-round leakage. Furthermore, our result determines the optimal mapping for the case with $Y_1 \leftrightarrow X_1 \leftrightarrow U$ and side information X_2 at agent 2 is also of independent interest.

2.3 Gaussian Sources: Interactive Mechanism

We now consider the case where the data pairs at each agent are drawn according to bivariate Gaussian distributions, i.e., $(X_1, Y_1) \sim N(0, \Sigma_{X_1,Y_1})$, $(X_2, Y_2) \sim N(0, \Sigma_{X_2,Y_2})$, and $(X_1, X_2) \sim N(0, \Sigma_{X_1,X_2})$. For jointly Gaussian sources subject to mean square error distortion constraints, we prove that one round of interaction suffices to achieve the utility-privacy tradeoff.

Theorem 4. *For the private interactive mechanism, the leakage-distortion region under mean square error distortion constraints consist of all tuples* (L_1, L_2, D_1, D_2) *satisfying*

$$L_1 \geq \frac{1}{2}\log(\frac{\sigma^2_{Y_1}}{\alpha^2 D_1 + \sigma^2_{Y_1|X_1,X_2}}) \tag{2.18}$$

$$L_2 \geq \frac{1}{2}\log(\frac{\sigma^2_{Y_2}}{\beta^2 D_2 + \sigma^2_{Y_2|X_1,X_2}}) \tag{2.19}$$

where $\alpha = \frac{cov(X_1,Y_1)}{\sigma^2_{Y_1}}$ *and* $\beta = \frac{cov(X_2,Y_2)}{\sigma^2_{Y_2}}$.

Proof. If (X_1, Y_1) is jointly Gaussian, we can write $Y_1 = \alpha X_1 + Z_1$, where Z_1 is a zero mean Gaussian random variable independent of X_1.

Achievability is established by considering Gaussian mechanism in each round, i.e., the sequence U_1^n is chosen such that the 'test channel' from U_1 to X_1 yields $U_1 = X_1 + V_1$, where V_1 is Gaussian and independent of the rest of random variables, with variance Q such that reconstruction function of $\hat{X}_1$ to be the MMSE estimate of X_1 given U_1 and X_2. For such a system, the minimum mean square error (MMSE) estimator minimizes the quadratic distortion measure. Therefore $D_1 = E(Var(X_1|U_1, X_2))$ (no interaction is required).

To prove the converse, we have

$$
\begin{aligned}
L_1 + \epsilon \geq & \frac{1}{n} I(Y_1^n; U_1^n, \ldots, U_K^n, X_2^n) && (2.20)\\
= & \frac{1}{n}[h(Y_1^n) - h(Y_1^n|U_1^n, \ldots, U_K^n, X_2^n)] && (2.21)\\
= & \frac{1}{n}[nh(Y_1) - \sum_{i=1}^{n} h(Y_{1i}|U_1^n, \ldots, U_K^n, X_2^n, Y_1^{i-1})] && (2.22)
\end{aligned}
$$

$$
\begin{aligned}
\geq & h(Y_1) - \frac{1}{n}\sum_{i=1}^{n} h(Y_{1i}|U_1^n, \ldots, U_K^n, X_2^n) && (2.23)\\
\geq & h(Y_1) - \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}\log(2\pi e(Var(Y_{1i}|U_1^n, \ldots, U_K^n, X_2^n))) && (2.24)\\
\geq & h(Y_1) - \frac{1}{2}\log(2\pi e\frac{1}{n}\sum_{i=1}^{n}(Var(Y_{1i}|U_1^n, \ldots, U_K^n, X_2^n))) && (2.25)\\
\geq & h(Y_1) - \frac{1}{2}\log(2\pi e\frac{1}{n}\sum_{i=1}^{n}(Var(\alpha X_{1i} + Z_{1i}|U_1^n, \ldots, U_K^n, X_2^n))) && (2.26)\\
\geq & \frac{1}{2}\log(\frac{\sigma_{Y_1}^2}{\alpha^2 D_1 + \sigma_{Y_1|X_1,X_2}^2}) && (2.27)
\end{aligned}
$$

Similarly, we can prove $L_2 \geq \frac{1}{2}\log(\frac{\sigma_{Y_2}^2}{\beta^2 D_2 + \sigma_{Y_2|X_1,X_2}^2})$. □

One can notice in the case that $Y_1 \leftrightarrow X_1 \leftrightarrow X_2 \leftrightarrow Y_2$ forms Markov chain we have $Var(Y_1|X_1, X_2) = Var(Y_1|X_1)$.

Chapter 3

PRIVATE INTERACTIVE MECHANISMS UNDER LOG-LOSS DISTORTION

Logarithmic loss is a widely used penalty function in machine learning theory and prediction and it is a natural loss criterion in scenarios where reconstructions are allowed to be soft, i.e., they can be probability measures instead of deterministic decision values. We now derive the leakage-distortion region under log-loss distortion.

Formally, for a random variable $X \in \mathcal{X}$ and its reproduction alphabet $\hat{\mathcal{X}}$ as the set of probability measures on $\mathcal{X}$, the log-loss distortion is defined as

$$d(x, \hat{x}) = \log(\frac{1}{\hat{x}(x)}). \tag{3.1}$$

3.0.1 Leakage-distortion region for log-loss distortion

Theorem 5. *For the K-round interaction mechanism the leakage-distortion region under log-loss distortion, set of all tuples (L_1, D_1, L_2, D_2) is given by:*

$$\begin{aligned} \{(L_1, L_2, D_1, D_2) : L_1 &\geq I(Y_1; U_1, \ldots, U_K, X_2), \\ L_2 &\geq I(Y_2; U_1, \ldots, U_K, X_1), \\ D_1 &\geq H(X_1 | U_1, \ldots, U_K, X_2) \\ D_2 &\geq H(X_2 | U_1, \ldots, U_K, X_1)\}. \end{aligned} \tag{3.2}$$

Proof. The distortion bounds in (3.2) result from applying $\hat{X}_i = P(X_i = x_i | U_1, \ldots, U_K, X_j)$

$i = 1, 2$, $j \neq i$, in Theorem 1, to get

$$D_i \geq E(d(X_i, \hat{X}_i))$$
$$= \sum_{x_i, u_1, \ldots, u_K} P(x_i, u_1, \ldots, u_K) \log\left(\frac{1}{P(x_i|u_1, \ldots, u_K, x_j)}\right) = H(X_i|U_1, \ldots, U_K, X_j), \tag{3.3}$$

where the summation is over $(x_i, u_1, \ldots, u_K)$ since $\hat{X}$ is a function of $(U_1, \ldots, U_K)$. □

Corollary 1. *For special case, $Y_i = X_i$, $i = 1, 2$, we have $L_1(D_1, D_2) = H(Y_1) - D_1$ and $L_2(D_1, D_2) = H(Y_2) - D_2$, i.e., the leakage for each agent is simply the rate-distortion function under log-loss distortion.*

For the case $Y_i = X_i$, $i = 1, 2$, as explained earlier, the leakage-distortion region is the same as the rate-distortion region. In [13], it is shown that a one-shot scheme achieves the rate-distortion region. In fact, the optimal mapping is a one-shot Wyner-Ziv scheme that each agent uses to share data simultaneously and independently with the other agent.

Corollary 2. *For $X_2 = Y_2 = \emptyset$, under log-loss distortion measure and $K = 1$, i.e., a a one-way (one round) single source agent and single receiver agent setting with no interaction (see Fig. 3.1) the bounds in Theorem 5 yields the following optimization problem.*

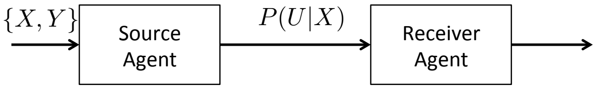

Figure 3.1: One-way non-interactive mechanism.

$$\min_{P(U|X): I(X;U) \geq \tau} I(Y;U). \tag{3.4}$$

In general, when $Y_i \neq X_i$, $i = 1, 2$, a one-shot scheme will not achieve the set of all (L_1, D_1, L_2, D_2) tuples in Theorem 5. It is then of interest to understand if interaction reduces leakage, and if so, what the optimal set of mechanisms are. To this end, we begin by rewriting the distortion bounds in (3.2) as

$$I(X_1; U_1, \ldots, U_K, X_2) \geq \tau_1 \tag{3.5}$$
$$I(X_2; U_1, \ldots, U_K, X_1) \geq \tau_2.$$

From (3.2), computing the K-round sum leakage leads to following optimization problem:

$$\min_{\{P_{1k}, P_{2k}\}_{k=1}^{K/2}} \sum_{i,j=1, i \neq j}^{2} I(Y_i; U_{1,\ldots,} U_K, X_j) \tag{3.6}$$

such that for all $i, j = 1, 2, i \neq j$,

$$I(X_i; U_1, \ldots U_K, X_j) \geq \tau_i. \tag{3.7}$$

The optimization problem (3.6) is not convex because of the non-convexity of the feasible region in (3.7). One can, however, draw parallels between the above optimization problem and the information bottleneck (IB) problem that Tishby *et al.* introduce in [14] in which for a source (X, Y) and an output U such that $Y \leftrightarrow X \leftrightarrow U$ form a Markov source, the goal is to minimize the information shared about X via U while preserving a measure of information about the correlated feature Y via U. One can see immediately that the IB problem is a dual of the privacy problem considered here in that the features to be revealed and hidden are swapped. Noting that the IB optimization problem is non-convex, the authors in [15] present an agglomerative information bottleneck algorithm that is guaranteed to converge to a local minima. Recently, in [6], Makhdoumi *et al.* also observe parallels between the information bottleneck problem and the single-round version of the problem considered here; i.e., for the case of a one-way non-interactive single source agent and a single receiver agent setup (with no side information at the receiver agent) shown in Fig. 3.1 and the associated "privacy funnel" optimization problem in (3.4). Furthermore, they apply Slonim's algorithm to their "privacy funnel" setup to compute a locally optimal mechanism. The optimization we study in (3.6) is an interactive version of (3.4), and thus, requires generalizing the methods and approaches for the non-interactive case to the interactive setup. In the following subsection we present an interactive version of the agglomerative IB algorithm and show how the presence of side-information in each round can be exploited to generalize the algorithm.

3.0.2 Information Bottleneck Problem and Agglomerative Information Bottleneck Algorithm

Consider the setting in Fig. 3.1 with $X_2 = \emptyset$ and $Y_2 = \emptyset$. The information bottleneck problem seeks to minimize the compression rate between X and U, while preserving a measure of the average information between U and some correlated data Y and is given by

$$\min_{P(U|X):I(Y;U)\geq\tau} I(X;U). \tag{3.8}$$

In [14], Tishby *et al.* showed that it is possible to characterize the general form of the locally optimal solution for the information bottleneck problem in (3.8). Tishby *et al.* also introduced an iterative algorithm that determines a locally optimal solution. A natural question to is whether one can extend the iterative algorithm to sum-leakage problem introduced in (3.6). An immediate obstacle is the fact the feasible region is not convex and distortion measures are not linear function of distributions. In spite of these drawbacks, in the following theorem we characterize the locally optimal solution of problem in (3.6). Note that one can also introduce an iterative algorithm to construct a locally optimal solution of (3.6) based on the following theorem.

Theorem 6. *Suppose that we are given condition in Theorem 1. The conditional distribution $P_{U_j|U^{j-1},X_1}(u_j|u^{j-1},x_1)$ for all j for Lagrange-multipliers β_1 and β_2 is the stationary point of Lagrangian*

$$\mathcal{L} = I(Y_1;U^K,X_2) + I(Y_2;U^K,X_1) - \beta_1 I(X_1;U^K,X_2) - \beta_2 I(X_2;U^K,X_1) \tag{3.9}$$

if and only if

$$P(u_j|u^{j-1},x_1) = \frac{P(u^j)}{\mathcal{Z}(x_1,x_2,u^{j-1},\beta_1,\beta_2)} \tag{3.10}$$

$$exp\Bigg\{ -\beta_1^{-1}[E_{X_2|X_1,u^{j-1}}\{D(P(y_1|x_1,x_2,u^{j-1})||P(y_1|u^j,x_2))\}$$
$$+D(P(y_2|x_1,u^{j-1})||P(y_2|x_1,u^j))] - D(P(x_2|x_1,u^{j-1})||P(x_2|u^j))\Bigg\} \tag{3.11}$$

for some β_1 and β_2, where $\mathcal{Z}(x_1,x_2,u^{j-1},\beta_1,\beta_2)$ is a normalization function.

Proof. The proof details can be found in Appendix C. □

Consider the mechanism depicted in Fig. 3.1. In this problem $X_2 = \emptyset$ and $Y_2 = \emptyset$. The optimal solution for this problem is given by

$$P(u|x) = \frac{P(u)}{\mathcal{Z}(x,\beta)} exp\{-\beta^{-1}D(P(y|x)||P(y|u)))\} \tag{3.12}$$

Several methods were developed to solve information bottleneck problem. However, for ease of computation and for cases in which a locally optimal solution may suffice, in [15], Slonim *et al.* propose an agglomerative algorithm which as the name suggests involves reducing the cardinality of U iteratively until the constraints on both X and Y are satisfied (since U acts as a quantized version of the feature to be minimally revealed) and prove that it converges to a local minima of the optimization problem. We adopt this algorithm and generalize it to the interactive setting.

We first briefly outline the agglomerative information bottleneck algorithm which yields a solution to (3.8). The procedure typically starts with the most fine-grained solution where $\mathcal{U} = \mathcal{X}$, i.e., each value of X is assigned to a unique singleton cluster in U. The idea is to reduce the cardinality of $\mathcal{U}$ and consequently reduce $I(X;U)$, by merging two values of $u_i \in \mathcal{U}$ and $u_j \in \mathcal{U}$ such that the new merged random variable U_{ij} is distributed as

$$P(u_{ij}|x) = P(u_i|x) + P(u_j|x) \tag{3.13}$$

In the k-th iteration, the indices i and j are chosen that U_{ij}^k satisfies the constraint in (3.8), while $I(X;U_{ij}^k)$ is at most as large as $I(X;U^{k-1})$ where U^{k-1} denotes the random variable from the previous iteration.

In[6], the authors apply the agglomerative information bottleneck algorithm to compute the locally optimal leakage for a desired τ in (3.4). They refer to the optimization problem in (3.4) as a privacy funnel problem and the resulting optimization algorithm as greedy algorithm privacy funnel.

As observed in [6], we note that the optimization problem in (3.6) as well as (3.4) differs from the information bottleneck problem in (3.8) in that the minimization and constraint functions are swapped for the same minimizing argument.

3.0.3 Agglomerative Interactive Privacy Algorithm

The optimization problem in (3.6) is an interactive generalization of the privacy funnel problem in (3.4) in which both agents have access to data sources that need to be shared. We now show that the multi-round interaction setup allows a natural generalization of the single round case. To develop such an algorithm, we first consider the single round case with side information at the receiver agent (depicted in Fig. 3.2). We introduce a *merge-and-search algorithm* that extends the agglomerative information bottleneck algorithm described earlier to a multivariate setting.

Merge-and-Search algorithm: Consider a one-round setting, i.e., $K = 1$ with side information at receiver agent (Fig. 3.2). Since $I(Y;Z)$ is fixed by joint source distribution, the optimization problem in (3.6) can be simplified as

$$\min_{P(U|X)} I(Y;U,Z) \text{ s.t. } I(X;U,Z) \geq \tau_1 \tag{3.14}$$

Comparing with (3.4), the optimization in (3.14) is obtained by replacing U by the tuple (U,Z) and $P(U|X)$ by $P(U,Z|X) = P(U|X)P(Z|X)$. Thus, in computing the optimal mechanism, one now needs to consider the pair (U,Z). We iteratively reduce the cardinality of U to reduce $I(Y;U,Z)$ by merging the values of U for each value of Z such that distortion condition in (3.14) is satisfied, i.e., in the k-th interaction, we choose indices i and j such that $I(X;U_{ij}^k,Z) \geq \tau_1$ where U_{ij}^k is the resulting from merging u_i and u_j while maximizing $I(Y;U^{k-1}|Z) - I(Y;U_{ij}^k|Z)$ where U^{k-1} is the output of the algorithm in round $k-1$. These steps are the basis of our merge-and-search algorithm that extends [6, Algorithm 1] to the more general point-to-point setting with side information at receiving agent.

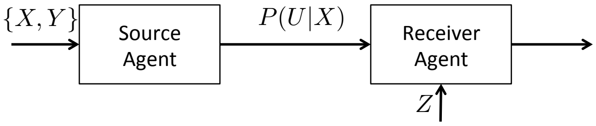

Figure 3.2: Point to point mechanism with side information

Consider the two-round setting in (3.6), i.e., $K = 2$. We can use the above described merge-and-search algorithm iteratively to find the mechanism (P_{11}, P_{21}). In the first round, we have a point-to-point setting with side information X_2 for which the distribution $P(U_1|X_1)$ can be found, as detailed above. In the second round, the cardinality of U_2 is reduced to decrease $I(Y_2;U_1,U_2,X_1)$ using $P(U_1,X_1)$ computed during the first round. This reduction is computed by merging elements of

Algorithm 1: Agglomerative Iterative Algorithm

```
For $k = 1, \ldots, K/2$
R(2k-1): $\min I(Y_1; X_2, U_1, \ldots, U_{2k-2}, U_{2k-1})$
        over $P(U_{2k-1}|X_2, U_1, \ldots, U_{2k-2})$
        s.t. $I(X_1; U_{2k_1}|X_2, U_1, \ldots, U_{2k-2}) \geq \tau_{2k-1}$
  Input (2k-1): $P(X_1, Y_1)$, $P(U_{2k-2}, \ldots, U_1, X_1, X_2)$, $\tau_{2k-1}$
    Apply the merge-and-search algorithm to find local optimum.
  Output (2k-1): $P(U_{2k-1}|X_1, X_2, U_1, \ldots, U_{2k-2})$
R(2k): $\min I(Y_2; X_1, U_1, \ldots, U_{2k-1}, U_{2k})$
        over $P(U_{2k}|X_1, U_1, \ldots, U_{2k-1})$
        s.t. $I(X_2; U_{2k}|X_1, U_1, \ldots, U_{2k-1}) \geq \tau_{2k}$
  Input (2k): $P(X_2, Y_2)$, $P(U_{2k-1}, \ldots, U_1, X_1, X_2)$, $\tau_{2k}$
    Apply the merge-and-search algorithm to find local optimum.
  Output (2k): $P(U_{2k}|X_1, X_2, U_1, \ldots, U_{2k-1})$
 Output : $P(U_1|X_1), \ldots, P(U_K|U_1, \ldots, U_{K-1}, X_2)$
```

U_2 conditioned on U_1 and X_1.

The steps we outlined above can be extended to find the locally optimal mechanism $\{P_{1i}, P_{2i}\}_{i=1}^{\frac{K}{2}}$ for any $K \geq 2$ and is detailed in Algorithm 1.

3.0.4 Gaussian Sources Under Log-Loss Distortion

In this section, we prove for Gaussian sources under log-loss distortion one round of interaction suffices. We leverage the results by Tishby *et.al* for the non-interactive case in [16] to prove that no interaction is required.

Proposition 1. *[[16], Theorem 1] Let (X, Y) be jointly Gaussian distributed. Let U be the output of a mapping $P(U|X)$ such that $Y \leftrightarrow X \leftrightarrow U$ forms a Markov chain. The mapping $P(U|X)$ that minimizes the following information bottleneck problem for jointly Gaussian sources*

$$\begin{aligned} \min_{P(U|X)} \quad & I(X;U) \\ \text{subject to} \quad & I(Y;U) \geq \tau. \end{aligned} \tag{3.15}$$

is Gaussian.

The above result extends in a straightforward manner to the non-interactive (one-way) privacy funnel setting given by (3.4) and we summarize it in the following corollary. The extension is a direct result of the fact that since X and Y are jointly Gaussian, the optimal mapping remains Gaussian even when the objective and constraint functions are swapped in (3.4).

Corollary 1. *For the non-interactive (one-way) single source and single receiver agent setting in Fig. 3.1 with the leakage-distortion tradeoff problem given by (3.4), the optimal leakage-minimizing mechanism is Gaussian.*

As a first step towards establishing optimality of a one-round Gaussian mechanism for the interactive setting, we extend the results in Proposition 1 to the case in which the receiver agent, chosen as agent 2 without loss of generality in the non-interactive setting, has side information Z correlated with source date (X, Y).

Lemma 3. *Suppose (X, Y) and (X, Z) are jointly Gaussian and let $P(U|X)$ be a privacy mechanism such that $U \leftrightarrow X \leftrightarrow Z$ forms a Markov chain (see Fig. 3.2). The optimal mechanism $P(U|X)$ minimizing $I(Y; U, Z)$ subject to $I(X; U, Z) \geq \tau$ is Gaussian.*

Proof. Define $V = (U, Z)$. Now, consider the following optimization problem

$$\begin{aligned} \min_{P(V|X)} \quad & I(Y; V) \\ \text{subject to} \quad & I(X; V) \geq \tau. \end{aligned} \tag{3.16}$$

.

From Corollary 1, the optimizing mechanism $P(V|X)$, and therefore, the output V in (3.16) are Gaussian. Thus, since Z is Gaussian, we have that (U, Z) are jointly Gaussian. Note that the mechanisms over which the optimizations are done in Lemma 3 and (3.16) are the same since $P(V|X) = P(U, Z|X) = P(Z|X)P(U|X)$ for a given source distribution $P(X, Z)$. □

We now use Lemma 3 to determine the optimal mechanism for the K-round interactive mechanism with Gaussian sources and show that one round of interaction suffices.

Theorem 7. *Consider a two-agent interactive setting with log-loss distortion and jointly Gaussian sources. The optimal leakage-distortion tradeoff region in Theorem 5 can be achieved in one round of interaction.*

Proof. From Lemma 3 we have that even with side information at the receiver agent, the optimal mechanism is Gaussian. Since the interactive setting involves a set of K such mechanisms, it is straightforward to see that the tuple $(U_1, \ldots, U_K)$ in Theorem 5 should also be Gaussian, i.e., one round of interaction suffices. □

3.0.5 Benefit of Interaction Under Log-Loss Distortion

In Theorem 5, we present the best achievable region under log-loss distortion. We now address the problem of whether more rounds can strictly improve the sum leakage-distortion function.

In this section, we show that there exists at least one source for which multiple rounds of interaction help under log-loss distortion. By using Theorem 3, we show interaction under log-loss distortion reduces leakage.

Theorem 8. *For a one-round interaction problem for a source* (X, Y) *at source agent with side information* Z *at receiver agent depicted in Fig. 3.2, there exists a joint probability distribution* $P_{X,Y,Z}$ *and distortion level* D *for which* $L^A_{sum,1}(P_{X,Y,Z}, D) > L^B_{sum,2}(P_{X,Y,Z}, D)$.

Proof. According to Theorem 3, it is sufficient to show there exist $P_{X,Y|Z}$ and distortion level D for which $\eta^A_1(P_{X,Y|Z}P_Z, D)$ is not a concave function with respect to P_Z. In particular, it is sufficient to show there exist P_{Z_1} and P_{Z_2} such that

$$\eta^A_1(P_{X,Y|Z}\frac{P_{Z_1} + P_{Z_2}}{2}, D) < \frac{\eta^A_1(P_{X,Y|Z}P_{Z_1}, D) + \eta^A_1(P_{X,Y|Z}P_{Z_2}, D)}{2} \tag{3.17}$$

Consider $\mathcal{X} = \mathcal{Y} = \{0, 1\}$. Let $P_{Z_1} \sim Bern(q)$ and$P_{Z_2} \sim Bern(\bar{q})$ where $\bar{q} = 1-q$. Let $P_{X,Y|Z}$ be the distribution in Table 3.1.

Table 3.1: Conditional Distribution $P_{X,Y|Z}$

$P_{X,Y\|Z}$	$Z = 0$	$Z = 1$
$X = 0, Y = 0$	$\bar{p}\bar{r}$	$p\bar{r}$
$X = 0, Y = 1$	$\bar{p}r$	pr
$X = 1, Y = 0$	pr	$\bar{p}r$
$X = 1, Y = 1$	$p\bar{r}$	$\bar{p}\bar{r}$

Let $P_Z = \frac{P_{Z_1}+P_{Z_2}}{2}$ which is $Bern(\frac{1}{2})$. The joint distribution (X, Z) is doubly symmetric binary source with parameter p, $(X, Z) \sim DSBS(p)$, and $Y = X + N$, $N \sim Bern(r)$ (with X and N independent). Exactly solving the maximization problem in both right and left-side of equation (3.17) is cumbersome but it is easy to provide bounds for them. Define $\phi(P_{X,Y,Z}, P_{U|X}) := H(Y|Z, U)$ and $\psi(P_{X,Y,Z}, P_{U|X}) := H(X|U, Z)$.

Lemma 4. *If $P_{Z_1} \sim Bern(q)$ and$P_{Z_2} \sim Bern(\bar{q})$ and $P_{X,Y|Z}$ is conditional distribution given in Table 3.1, then*

$$\frac{\eta_1^A(P_{X,Y|Z}P_{Z_1}, D) + \eta_1^A(P_{X,Y|Z}P_{Z_2}, D)}{2} \geq C(p, q, r, \alpha_{2,0}, \alpha_{2,1}) \tag{3.18}$$

holds for

$$D = \gamma(p, q, r, \alpha_{2,0}, \alpha_{2,1}) \tag{3.19}$$

where

$$\begin{aligned} C(p, q, r, \alpha_{2,0}, \alpha_{2,1}) = \phi(P_{X,Y|Z}P_{Z_1}, \alpha_{2,0}, \alpha_{2,1}) =& (\bar{p}\bar{q}\alpha_{2,0} + \bar{p}q\alpha_{2,1})H(\frac{\bar{p}\bar{r}\alpha_{2,0} + pr\alpha_{2,1}}{\bar{p}\alpha_{2,0} + p\alpha_{2,1}}) \\ &+ (pq\alpha_{2,0} + \bar{p}q\alpha_{2,1})H(\frac{p\bar{r}\alpha_{2,0} + \bar{p}r\alpha_{2,1}}{p\alpha_{2,0} + \bar{p}\alpha_{2,1}}) \end{aligned} \tag{3.20}$$

and

$$\begin{aligned} \gamma(p, q, r, \alpha_{2,0}, \alpha_{2,1}) = \psi(P_{X,Y|Z}P_{Z_1}, \alpha_{2,0}, \alpha_{2,1}) =& \bar{q}(\bar{p}\alpha_{2,0} + p\alpha_{2,1})H(\frac{\bar{p}\alpha_{2,0}}{\bar{p}\alpha_{2,0} + p\alpha_{2,1}}) \\ & q(p\alpha_{2,0} + \bar{p}\alpha_{2,1})H(\frac{p\alpha_{2,0}}{p\alpha_{2,0} + \bar{p}\alpha_{2,1}}) \end{aligned} \tag{3.21}$$

where $H(.)$ is binary entropy and distribution

$$P(U|X) = \begin{bmatrix} 1 - \alpha_{2,0} & 0 \\ 0 & 1 - \alpha_{2,1} \\ \alpha_{2,0} & \alpha_{2,1} \end{bmatrix}$$

where $0 \leq \alpha_{2,0}, \alpha_{2,1} \leq 1$.

Proof. Note that $\eta_1^A(P_{X,Y,Z_1}, D)$ is given by

$$\begin{aligned} \max_{P(U|X)} \quad & H(Y|Z_1, U) \\ \text{subject to} \quad & H(X|Z_1, U) \leq D \end{aligned} \tag{3.22}$$

which is greater than or equal to the objective value of the solution of following problem.

$$\begin{aligned} \max_{P(U|X)} \quad & H(Y|Z_1,U) \\ \text{subject to} \quad & H(X|Z_1,U) \leq D \\ & P(U|X) = \begin{bmatrix} 1-\alpha_{2,0} & 0 \\ 0 & 1-\alpha_{2,1} \\ \alpha_{2,0} & \alpha_{2,1} \end{bmatrix}. \end{aligned} \tag{3.23}$$

Now, by substituting distribution $P(U|X)$ and computing $\phi(P_{X,Y|Z}P_{Z_1},\alpha_{2,0},\alpha_{2,1})$, it can be seen that (3.23) is greater than or equal to $C(p,q,r,\alpha_{2,0},\alpha_{2,1})$. Consequently, we have $\eta_1^A(P_{X,Y|Z}P_{Z_1},D) \geq C(p,q,r,\alpha_{2,0},\alpha_{2,1})$. Observe that $C(p,q,r,\alpha_{2,0},\alpha_{2,1}) = C(p,\bar{q},r,\alpha_{2,0},\alpha_{2,1})$ and $\gamma(p,q,r,\alpha_{2,0},\alpha_{2,1}) = \gamma(p,\bar{q},r,\alpha_{2,0},\alpha_{2,1})$ hold. Therefore we have

$$\begin{aligned} \eta_1^A(P_{X,Y|Z}P_{Z_2},D) \geq & C(p,\bar{q},r,\alpha_{2,0},\alpha_{2,1}) \\ = & C(p,q,r,\alpha_{2,0},\alpha_{2,1}) \end{aligned} \tag{3.24}$$

it follows that

$$\frac{\eta_1^A(P_{X,Y|Z}P_{Z_1},D)+\eta_1^A(P_{X,Y|Z}P_{Z_2},D)}{2} \geq C(p,q,r,\alpha_{2,0},\alpha_{2,1}) \tag{3.25}$$

□

Left-side of (3.17) is given by

$$\begin{aligned} \max_{P(U|X)} \quad & H(Y|Z,U) \\ \text{subject to} \quad & H(X|Z,U) \leq D. \end{aligned} \tag{3.26}$$

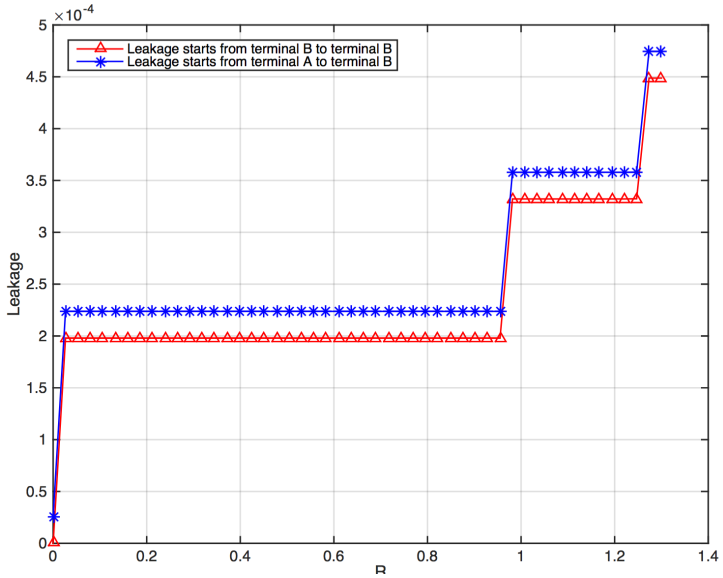

Figure 3.3: Comparing Sum leakage for the two round vs the one round interactive mechanism, where the blue curve with stars and the red curve with triangles show the one-round and the two-round interaction mechanism, respectively.

Equation (3.26) is equivalent to

$$\begin{aligned} \max_{P(U|X)} \quad & H(X|Z,U) + H(Y|X,Z) - H(X|Y,Z,U) \\ \text{subject to} \quad & H(X|Z,U) \le D. \end{aligned} \tag{3.27}$$

Which is less than or equal to $D + H(r)$. For all $q \in (0, \frac{1}{2})$ and all $\alpha_{2,0}, \alpha_{2,1} \in (0,1)$, there exists an r such that $D + H(r)$ is strictly less than $C(p, q, r, \alpha_{2,0}, \alpha_{2,1})$. □

3.1 Illustration of results

We illustrate our results for the log-loss distortion measure, and in particular, explore the effect of interaction on leakage using a publicly available dataset. The US Census dataset is a sample of US population from 1994. It contains different features including age, ethnicity, income levels, work class, and, gender such that the age feature is categorized into 7 levels, gender and income level (above 50K USD and less than 50K USD) are binary random variables, Work class is categorized in 4 levels, and, ethnicity is classified into 4 levels. We choose X_1 =(age, gender), X_2 =

(ethnicity, gender), Y_1 =(work class), and, Y_2 =(income level), thus wishing to keep private work class and income level at agents 1 and 2, respectively.

In Fig. 3.3, using Algorithm 1 and the empirical distribution of the data, we plot both the one-round and the two-round sum leakages as functions of mutual information based on log-loss distortion level at agent B. To demonstrate the value of interaction we consider the following results: let $d_A = 0$ and d_B be the log-loss distortion measure. The blue curve with stars is the leakage for one round from A to B. We note that it upper bounds the red curve with triangles which denotes the sum leakage starting from B to A and back to B, thus suggesting the interaction can reduce leakage for the log-loss setting.

Chapter 4

CONCLUSION AND FUTURE WORK

We have defined a K-round interactive privacy mechanism between two agents with correlated sources, and have determined the leakage-distortion region for general distortion functions with particular focus on log-loss distortion. For both general and log-loss distortion functions, we have illustrated that interaction can reduce leakage. In practice, this suggests that agents can share just sufficient data to achieve distortion over multiple results relative to a one-shot non-interactive setting. Future work includes evaluating leakage for different classes of statistical inference attacks as well as extension to the multi-agent ($K > 2$) case.

REFERENCES

[1] C. Dwork, “Differential privacy,” in *Automata, languages and programming*, pp. 1–12, Springer, 2006.

[2] 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, pp. 838–852, June 2013.

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

[4] S. Salamatian, A. Zhang, F. Calmon, S. Bhamidipati, N. Fawaz, B. Kveton, P. Oliveira, and N. Taft, “How to hide the elephant- or the donkey- in the room: Practical privacy against statistical inference for large data,” *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, 2013.

[5] L. Sankar, S. Kar, R. Tandon, and H. V. Poor, “Competitive privacy in the smart grid: An information-theoretic approach,” in *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, pp. 220–225, IEEE, 2011.

[6] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Medard, “From the information bottleneck to the privacy funnel,” in *Information Theory Workshop (ITW), 2014 IEEE*, pp. 501–505, Nov 2014.

[7] M. Vera, L. Vega, and P. Piantanida, “The two-way cooperative information bottleneck,” *IEEE, International Symposium on Information Theory(ISIT)*, 2015.

[8] S. Kairouz, P. Oh and P. Viswanath, “Optimalitlity of non-interactive randomized response,” *CoRR*, vol. abs/1407.1546, 2014.

[9] A. Kaspi, “Two-way source coding with a fidelity criterion,” *Information Theory, IEEE Transactions on*, vol. 31, pp. 735–740, Nov 1985.

[10] N. Ma and P. Ishwar, “The infinite-message limit of two-terminal interactive source coding,” *Information Theory, IEEE Transactions on*, vol. 59, pp. 4071–4094, July 2013.

[11] N. Ma, P. Ishwar, and P. Gupta, “Interactive source coding for function computation in collocated networks,” *Information Theory, IEEE Transactions on*, vol. 58, no. 7, pp. 4289–4305, 2012.

[12] A. El Gamal and Y. Kim, *Network information theory.* Cambridge university press, 2011.

[13] T. Courtade and R. Wesel, “Multiterminal source coding with an entropy-based distortion measure,” in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pp. 2040–2044, July 2011.

[14] N. Tishby, F. Pereira, and W. Bialek, “The information bottleneck method,” *DBLP: journals/corr/physics-004057*, 2000.

[15] N. Slonim and N. Tishby, “Agglomerative information bottleneck,” *Proc. of Neural Inforamtion Processing System (NIPS-99)*, 1999.

[16] G. Chechik, A. Globerson, N. Tishby, and Y. Weiss, “Information bottleneck for gaussian variables,” *in Journal of Machine Learning Research*, 2004.

[17] H. Yamamoto, “A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers (corresp.),” *Information Theory, IEEE Transactions*, vol. 29, pp. 918–923, Nov 1983.

APPENDIX A

PROOF OF THEOREM 1

Proof. Achievability: we will show that following inequality holds.

$$\lim_{n\to\infty}\frac{1}{n}I(Y_1^n;U_1^n,\ldots,U_K^n,X_2^n)\leq I(Y_1;\ U_1,\ldots,U_K,X_2)+\epsilon \tag{A.1}$$

$$I(Y_1^n;U_1^n,\ldots,U_K^n,X_2^n)=H(Y_1^n)-H(Y_1^n|U_1^n,\ldots,U_K^n,X_2^n) \tag{A.2}$$

$$=nH(Y_1)-H(Y_1^n|U_1^n,\ldots,U_K^n,X_2^n) \tag{A.3}$$

$$=nH(Y_1)-\sum_{u_1,\ldots,u_K,x_2}P(u_1,\ldots,u_K,x_2)H(Y_1^n|u_1,\ldots,u_K,x_2) \tag{A.4}$$

$$=nH(Y_1)-[\sum_{u_1,\ldots,u_K,x_2\in\mathcal{T}_{U_1,\ldots,U_K,X_2}}P(u_1,\ldots,u_K,x_2)H(Y_1^n|u_1,\ldots,u_K,x_2)$$

$$+\sum_{u_1,\ldots,u_K,x_2\notin\mathcal{T}_{U_1,\ldots,U_K,X_2}}P(u_1,\ldots,u_K,x_2)H(Y_1^n|u_1,\ldots,u_K,x_2)] \tag{A.5}$$

$$=nH(Y_1)-[\sum_{u_1,\ldots,u_K,x_2\in\mathcal{T}_{U_1,\ldots,U_K,X_2}}P(u_1,\ldots,u_K,x_2)\{$$

$$-\sum_{y_1\in\mathcal{T}_{Y_1|u_1,\ldots,u_K,x_2}}P(y_1|u_1,\ldots,u_K,x_2)\log(P(y_1|u_1,\ldots,u_K,x_2))$$

$$-\sum_{y_1\notin\mathcal{T}_{Y_1|u_1,\ldots,u_K,x_2}}P(y_1|u_1,\ldots,u_K,x_2)\log(P(y_1|u_1,\ldots,u_K,x_2))\}]$$

$$-\sum_{u_1,\ldots,u_K,x_2\notin\mathcal{T}_{U_1,\ldots,U_K,X_2}}P(u_1,\ldots,u_K,x_2)H(Y_1^n|u_1,\ldots,u_K,x_2)] \tag{A.6}$$

$$=nH(Y_1)-nH(Y_1|U_1,\ldots,U_K,X_2)$$

$$+\sum_{u_1,\ldots,u_K,x_2\in\mathcal{T}_{U_1,\ldots,U_K,X_2}}P(u_1,\ldots,u_K,x_2)\{$$

$$-\sum_{y_1\notin\mathcal{T}_{Y_1|u_1,\ldots,u_K,x_2}}P(y_1|u_1,\ldots,u_K,x_2)\log(P(y_1|u_1,\ldots,u_K,x_2))\}] \tag{A.7}$$

$$-\sum_{u_1,\ldots,u_K,x_2\notin\mathcal{T}_{U_1,\ldots,U_K,X_2}}P(u_1,\ldots,u_K,x_2)H(Y_1^n|u_1,\ldots,u_K,x_2)] \tag{A.8}$$

$$\leq nI(Y_1;U_1,\ldots,U_K,X_2)+n\epsilon(n) \tag{A.9}$$

Where $\mathcal{T}_{U_1,\ldots,U_K,X_2}$, $\mathcal{T}_{Y_1|u_1,\ldots,u_K,x_2}$ are sets of jointly typical sequences and conditional typical sequences, respectively, and (A.2) follows from definition of mutual information, (A.4) is definition of conditional entropy and (A.6)-(A.8) follows from this fact that the probability of being non-typical sequence goes to zero as n goes to infinity. Note that $\epsilon(n)$ goes to zero when n goes to infinity.

Converse: To prove the converse, according to (2.3a)-(2.3d) we are given a mechanism, $U_1^n, \ldots, U_K^n$. (2.3c) implies

$$L_1 + \epsilon \geq \frac{1}{n} I(Y_1^n; U_1^n \ldots, U_K^n, X_2^n) \tag{A.10}$$

$$= \frac{1}{n}[I(Y_1^n; X_2^n) + I(Y_1^n; U_1^n \ldots, U_K^n | X_2^n)] \tag{A.11}$$

$$= \frac{1}{n} \sum_{i=1}^{n} [I(Y_{1i}^n; X_{2i}^n)$$

$$+ H(Y_{1i}|X_{2i}) - H(Y_{1i}|U_1^n \ldots, U_K^n, X_2^n, Y_1^{i-1})] \tag{A.12}$$

$$\geq \frac{1}{n} \sum_{i=1}^{n} [I(Y_{1i}^n; X_{2i}^n)$$

$$+ H(Y_{1i}|X_{2i}) - H(Y_{1i}|U_1^n \ldots, U_K^n, X_{2i}^n)] \tag{A.13}$$

$$= \frac{1}{n} \sum_{i=1}^{n} [I(Y_{1i}^n; X_{2i}^n) + I(Y_{1i}; U_1^n \ldots, U_K^n | X_{2i})] \tag{A.14}$$

$$= \frac{1}{n} \sum_{i=1}^{n} [I(Y_{1i}^n; X_{2i}^n) + I(Y_{1i}; U_{1i} \ldots, U_{Ki} | X_{2i})] \tag{A.15}$$

$$= \frac{1}{n} \sum_{i=1}^{n} I(Y_{1i}; U_{1i} \ldots, U_{Ki}, X_{2i}) \tag{A.16}$$

where (A.10) follows from (2.3c), (A.11) is definition of mutual information, (A.12) follows from chain rule and the fact that sources are i.i.d., (A.13) follows from this fact that condition reduces entropy. Note that leakage-distortion function is non-increasing and convex function of D[17].

We now show that following Markov chains holds

$$Y_{1i} \leftrightarrow (U_{1i}, \ldots, U_{2k-1,i}, X_{2i}) \leftrightarrow U_{2k,i} \tag{A.17}$$

$$Y_{2i} \leftrightarrow (U_{1i}, \ldots, U_{2k-2,i}, X_{1i}) \leftrightarrow U_{2k-1,i} \tag{A.18}$$

(A.17) holds because our sources are i.i.d. random variables and according to our mechanism $U_{2k,i}$ is independent of Y_{1i} condition on $(U_{1i}, \ldots, U_{2k-1,i}, X_{2i})$. Similarly, we can show that (A.18) holds. □

APPENDIX B

PROOF OF (10)

Proof. From (2.8), we have

$$L^A_{sum,1}(0, D_2) = \min_{P_{U_1|X_1}} [I(X_1; Y_2) + I(Y_1; U_1, X_2)] \tag{B.1}$$

For $d_A = 0$ and d_B in (2.15) with distortion level D_2, $L^A_{sum,1}(0, D_2) = 2 - H(2p(1 - p)) - \max_{P(U_1|X_1)} H(Y_1|U_1, X_2)$ where $\mathcal{U} = \{0, e, 1\}$ and

$$P(U_1|X_1) = \begin{cases} \alpha_0, & \text{if } x = 0 \text{ and } u = e \\ 1 - \alpha_0, & \text{if } x = 0 \text{ and } u = 0 \\ \alpha_1, & \text{if } x = 1 \text{ and } u = e \\ 1 - \alpha_1, & \text{if } x = 1 \text{ and } u = 1 \\ 0, & \text{otherwise} \end{cases} \tag{B.2}$$

where $E(d_B(X_1, U_1)) = P_{X_1}(0)\alpha_0 + P_{X_1}(1)\alpha_1 \leq D_2$. We have $P(X_1 = 0, U_1 = 1) = P(X_1 = 1, U_1 = 0) = 0$ because otherwise $E(d_B(X_1, U_1)) = \infty$. Thus, we have

$$H(Y_1|U_1, X_2) = \frac{1}{2}(1 - \alpha_0)H(p) + \frac{1}{2}(1 - \alpha_1)H(p) \tag{B.3}$$

$$+[\frac{\alpha_0}{2}(1 - p) + \frac{\alpha_1}{2}p]H(\frac{(1 - p)^2\alpha_0 + p^2\alpha_1}{(1 - p)\alpha_0 + p\alpha_1}) \tag{B.4}$$

$$+[\frac{\alpha_0}{2}p + \frac{\alpha_1}{2}(1 - p)]H(\frac{p(1 - p)\alpha_0 + p(1 - p)\alpha_1}{p\alpha_0 + (1 - p)\alpha_1}) \tag{B.5}$$

$H(Y_1|U_1, X_2)$ is maximized if $\alpha_0 = \alpha_1 = \alpha$, then the result is attained. □

APPENDIX C

PROOF OF THEOREM 6

We need to consider

$$\begin{aligned}\mathcal{L}' =& I(Y_1;U^K,X_2)+I(Y_2;U^K,X_1)-\beta_1 I(X_1;U^K,X_2)-\beta_2 I(X_2;U^K,X_1)\\ &+\sum_{x_1,x_2}\lambda(x_1,x_2)\sum_{u^K}P(u^K|x_1,x_2)\end{aligned} \tag{C.1}$$

where the last term corresponds to the normalization constraint. By expanding $\mathcal{L}'$ we have

$$\begin{aligned}\mathcal{L}' =& \sum_{y_1,u^K,x_2}P(y_1,u^K,x_2)log\frac{P(y_1,u^K,x_2)}{P(y_1)P(u^K,x_2)}+\sum_{y_2,u^K,x_1}P(y_2,u^K,x_1)log\frac{P(y_2,u^K,x_1)}{P(y_2)P(u^K,x_1)}\\ &-\beta_1\sum_{x_1,u^K,x_2}P(x_1,u^K,x_2)log\frac{P(x_1,u^K,x_2)}{P(x_1)P(u^K,x_2)}-\beta_2\sum_{x_1,u^K,x_2}P(x_1,u^K,x_2)log\frac{P(x_1,u^K,x_2)}{P(x_2)P(u^K,x_1)}\\ &+\sum_{x_1,x_2}\lambda(x_1,x_2)\sum_{u^K}P(u^K|x_1,x_2)\end{aligned} \tag{C.2}$$

By differentiating with respect to $P(u_j|u^{j-1},x_1)$ we have the following

$$\frac{\partial P(y_1,u^K,x_2)}{\partial P(u_j|u^{j-1},x_1)}=P(x_1,x_2,y_1,u^K_{-j}) \tag{C.3}$$

$$\frac{\partial P(y_2,u^K,x_1)}{\partial P(u_j|u^{j-1},x_1)}=P(x_2,y_1,u^K_{-j}) \tag{C.4}$$

$$\frac{\partial P(x_1,u^K,x_2)}{\partial P(u_j|u^{j-1},x_1)}=P(x_1,x_2,u^K_{-j}) \tag{C.5}$$

$$\frac{\partial P(u^K,x_2)}{\partial P(u_j|u^{j-1},x_1)}=P(x_1,x_2,u^K_{-j}) \tag{C.6}$$

$$\frac{\partial P(u^K,x_1)}{\partial P(u_j|u^{j-1},x_1)}=P(x_1,u^K_{-j}) \tag{C.7}$$

where u_{-j}^K is $u^K - \{u_j\}$. By using above partial derivative, we now differentiate $\mathcal{L}'$ with respect to $P(u_j|u^{j-1}, x_1)$.

$$\frac{\partial \mathcal{L}'}{\partial P(u_j|u^{j-1}, x_1)} = \sum_{y_1, x_2, u_{j+1}^K} P(x_1, y_1, x_2, u^K{-}j) \log(P(y_1|x_2, u^K))$$

$$+ \sum_{y_1, u_{j+1}^K} P(x_1, y_2, u^K{-}j) \log(P(y_2|x_1, u^K)) - \beta_1 \sum_{x_2, u_{j+1}^K} P(x_1, x_2, u^K{-}j) \log(\frac{P(x_1, x_2, u^K)}{P(x_2, u^K)})$$

$$- \beta_2 \sum_{x_2, u_{j+1}^K} P(x_1, x_2, u^K{-}j) \log(\frac{P(x_1, x_2, u^K)}{P(x_1, u^K)}) + \bar{\lambda}(x_1, x_2) \tag{C.8}$$

Dividing equations in (C.8) by $P(x_1, u^{j-1})$ and rearranging and putting equal to zero we have

$$- E_{X_2|X_1, U^{j-1}}(D(P(y_1|x_1, x_2, U^{j-1})||P(y_1|u^j, x_2))) - D(P(y_2|x_1, U^{j-1})||P(y_2|x_1, u^j))$$

$$- \beta_1 \log(P(u_j|u^{j-1}, x_1)) - \beta_1 D(P(x_2|x_1, U^{j-1})||P(x_2|u^j)) + \beta_1 \log(P(u^j)) + \hat{\lambda}(x_1, x_2, u^{j-1}, \beta_1, \beta_2) = \tag{C.9}$$

where $D(.||.)$ is KL distance and we absorb all the terms that don't depend on U_j in $\hat{\lambda}$.