Data Sanitisation Protocols for the Privacy Funnel with Differential Privacy Guarantees

Milan Lopuhaä-Zwakenberg, Haochen Tong, and Boris Škorić
Department of Mathematics and Computer Science
Eindhoven University of Technology
Eindhoven, the Netherlands
email: {m.a.lopuhaa,b.skoric}@tue.nl, h.tong@student.tue.nl

Abstract—**In the Open Data approach, governments and other public organisations want to share their datasets with the public, for accountability and to support participation. Data must be opened in such a way that individual privacy is safeguarded. The Privacy Funnel is a mathematical approach that produces a sanitised database that does not leak private data beyond a chosen threshold. The downsides to this approach are that it does not give worst-case privacy guarantees, and that finding optimal sanitisation protocols can be computationally prohibitive. These problems are tackled by using differential privacy metrics, and by considering local protocols that operate on one entry at a time. It is shown that under both the Local Differential Privacy and Local Information Privacy leakage metrics, one can efficiently obtain optimal protocols. Furthermore, Local Information Privacy is more closely aligned to the privacy requirements of the Privacy Funnel scenario, and optimal protocols satisfying Local Information Privacy are more efficiently computable. This paper also considers the scenario where each user has multiple attributes, for which a side-channel resistant privacy criterion is defined, and efficient methods to find protocols satisfying this criterion, while still offering good utility, are given. Finally, Conditional Reporting is introduced, an explicit LIP protocol that can be used when the optimal protocol is infeasible to compute. Experiments on real-world and synthetic data confirm the validity of these methods. The main output of this paper consists of methods to compute optimal privacy protocols, and explicit privacy protocols when the former are unfeasible computationally.**

***Keywords*—Privacy funnel; local differential privacy; information privacy; database sanitisation; complexity.**

I. INTRODUCTION

This paper is an extended version of [1]. Under the Open Data paradigm, governments and other public organisations want to share their collected data with the general public. This increases a government's transparency, and it also gives citizens and businesses the means to participate in decision-making, as well as using the data for their own purposes. However, while the released data should be as faithful to the raw data as possible, individual citizens' private data should not be compromised by such data publication.

Let $\mathcal{X}$ be a finite set. Consider a database $\vec{X} = (X_1, \dots, X_n) \in \mathcal{X}^n$ owned by a data aggregator, containing a data item $X_i \in \mathcal{X}$ for each user i (For typical database settings, each user's data is a vector of attributes $X_i = (X_i^1, \dots, X_i^m)$; this will be considered in more detail in Section VI). This data may not be considered sensitive by itself, but it might be correlated to a secret S_i. For instance, X_i might contain the age, sex, weight, skin colour, and average blood pressure of person i, while S_i is the presence of some medical condition. To publish the data in a privacy-preserving manner, the aggregator releases a sanitised database $\vec{Y} = (Y_1, \dots, Y_n)$, obtained from applying a sanitisation mechanism $\mathcal{R}$ to $\vec{X}$. In this setting, *privacy* is considered to be the extent to which one is unable to infer information about the S_i from the sanitised database $\vec{Y}$. One way to formulate this is by measuring the privacy leakage as the mutual information $\mathrm{I}(\vec{S};\vec{Y})$, and utility as the mutual information $\mathrm{I}(\vec{X};\vec{Y})$. This leads to the *Privacy Funnel* (PF) problem:

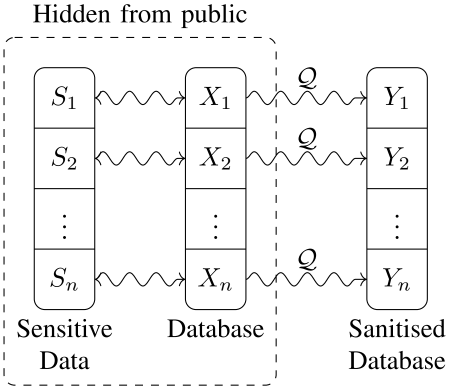


Figure 1. Model of PF with local protocols.

Problem 1. (Privacy Funnel, [2]) *Suppose the joint probability distribution of $\vec{S}$ and $\vec{X}$ is known to the aggregator, and let $M \in \mathbb{R}_{\geq 0}$. Then, find the sanitisation mechanism $\mathcal{R}$ such that $\mathrm{I}(\vec{X};\vec{Y})$ is maximised while $\mathrm{I}(\vec{S};\vec{Y}) \leq M$.*

There are two difficulties with this approach:

1) Finding and implementing good sanitisation mechanisms that operate on all of $\vec{X}$ can be computationally prohibitive for large n, as the complexity is exponential in n [3][4].
2) Taking mutual information as a leakage measure has as a disadvantage that it gives guarantees about the leakage in the average case. If n is large, this still leaves room for the sanitisation protocol to leak undesirably much information about a few unlucky users.

To deal with these two difficulties, two changes are made to the general approach. First, the focus is on *local* data sanitisation, i.e., optimisation protocols $\mathcal{Q}\colon \mathcal{X} \to \mathcal{Y}$ are considered, for some finite set $\mathcal{Y}$, and $\mathcal{Q}$ is applied to each X_i individually; this situation is depicted in Figure 1. Local sanitisation can be implemented efficiently. In fact, this approach is often taken in the PF setting [5][3]. Second, to ensure strong privacy guarantees even in worst-case scenarios,stricter notions of privacy are considered, based on Local Differential Privacy (LDP) [6]. For these metrics, methods are developed to find optimal protocols. Furthermore, for situations where the optimal protocol is computationally unfeasible to find, a new protocol is introduced, *Conditional Reporting* (CR), that takes advantage of the fact that only S_i needs to be protected. Determining CR only requires finding the root of a onedimensional increasing function, which can be done fast numerically.

A. New contributions

In this paper, two Differential Privacy-like privacy metrics are adapted to the PF situation, namely ε-LDP [6] and Local Information Privacy (ε-LIP) [7][8]. These metrics are modified so that they measure leakage about the underlying S rather than X itself (for notational convenience, S, X, Y rather than S_i, X_i, Y_i is used throughout the rest of this paper). For a given level of leakage, the aim is to find the privacy protocol that maximises the mutual information between input X_i and output Y_i. Adapting methods from [9] on LDP and [10] on perfect privacy, the following Theorem is proven:

Theorem 1 (Theorems 2 and 3 paraphrased)**.** *Suppose X and S are discrete random variables on sets of size a and c, respectively. Suppose that their joint distribution and a privacy level $\varepsilon \geq 0$ are given.*

1) *The optimal ε-LDP protocol can be found by enumerating the vertices of a polytope in $a^2 - a$ dimensions defined by $a(c^2 - c)$ inequalities.*
2) *The optimal ε-LIP protocol can be found by enumerating the vertices of a polytope in $a - 1$ dimensions defined by $2ac$ inequalities.*

This theorem gives us methods to get data sanitisation protocols that give strong privacy guarantees, and optimal utility under these guarantees. This is important in settings where worst-case guarantees for privacy leakage are needed, rather than a bound on the average user's privacy.

Since the complexity of the polytope vertex enumeration depends significantly on both its dimension and the number of defining inequalities [11], finding optimal LIP protocols can be done significantly faster than finding optimal LDP protocols. Furthermore, it will be argued that LIP is a privacy metric that more accurately captures information leakage than LDP in the PF scenario. For these two reasons only LIP is considered in the remainder of the paper, although many results can also be formulated for LDP.

A common scenario is that a user's data X consists of multiple attributes, i.e., $X = (X^1, \ldots, X^m)$. Here one can consider an attacker model where the attacker has access to some of the X^j. In this situation ε-LIP does not accurately reflect a user's privacy. Because of this, a new privacy condition called *Side-channel Resistant LIP* is introduced that takes such sidechannels into account, and methods to find optimal protocols that satisfy this privacy condition are described.

Finding the optimal protocols can become computationally unfeasible for large a and c. In such a situation, one needs to resort to explicitely given protocols. In the literature there is a wealth of protocols that satisfy ε-LDP w.r.t. X. These certainly work in the PF situation, but they might not be ideal, because these are designed to obfuscate all information about X, rather than just the part that relates to S. For this reason, Conditional Reporting (CR) is introduced, a privacy protocol that focuses on hiding S rather than X. Finding the appropriate CR protocol for a given probability distribution and privacy level can be done fast numerically.

The structure of this paper is as follows. In Section II, an overview is given of related work on PF, LDP, and finding optimal protocols. The mathematical setting of this paper is formalised in Section III. In Sections IV and V, Theorem 1 is proven for LDP and LIP, respectively. In Section VI privacy in the multiple attribute scenario is discussed. Section VII is dedicated to Conditional Reporting and its privacy properties. In Section VIII, he methods and protocols discussed above are tested on both synthetic and real data. Compared to [1], new contents in this extended paper are Section VII, the experiments on real data, and the extended literature review.

II. Related work

The PF setting was introduced in [5], to provide a framework for obfuscating data in such a way that the obfuscated data remains as faithful as possible to the original, while ensuring that the information leakage about a latent variable is limited. PF is related to the Information Bottleneck (IB) [12], a problem from machine learning that seeks to compress data as much as possible, while retaining a minimal threshold of information about a latent variable. In PF as well as IB, both utility and leakage are measured via mutual information. Many approaches to finding the optimal protocols in PF also work for IB and vice versa [13][3]. A wider range of privacy metrics for PF, and their relation to Differential Privacy, is discussed in [8].

LDP was introduced in [6]. It is an adaptation of Differential Privacy (DP) [14] to a setting where there is no trusted central party to obfuscate the data. As a privacy metric, it has the advantage that it offers a privacy guarantee in any case, not just the average case, and that it does not depend on the data distribution. On the downside, it can be difficult to fulfill such a stringent definition of privacy, and many relaxations of (L)DP have been proposed [15][16][17][18]. Of particular interest to this paper is LIP [7][8], also called Removal Local Differential Privacy [19]. LIP retains the worst-case guarantees of LDP, but is less restrictive, and can take advantage of a known distribution. In the context where only part of the data is considered secret, many privacy metrics fall under the umbrella of Pufferfish Privacy [20].

In [9], a method was introduced for finding optimal LDP-protocols for a wide variety of utility metrics, including mutual information. The method relies on finding the vertices of a polytope, but since this is the well-studied Differential Privacy polytope, its vertices can be described explicitly [21]. Similarly, [10] uses a vertex enumeration method to find the optimal protocol in the perfect privacy situation, i.e., when the released data is independent of the secret data. The complexity of vertex enumeration is discussed in [22][11].

One can conclude that PF and LDP are both well-studied, and so are methods to find optimal LDP protocols. However, LDP-like metrics so far have not been applied to the PF scenario. The aim of this paper is to do so, and to find optimal PF protocols that satisfy LDP-like privacy requirements.

III. Mathematical Setting

The database $\vec{X} = (X_1, \dots, X_n)$ consists of a data item X_i for each user i, each an element of a given finite set $\mathcal{X}$. Furthermore, each user has sensitive data $S_i \in \mathcal{S}$, which is correlated with X_i; again $\mathcal{S}$ is assumed to be finite (see Figure 1). Each (S_i, X_i) is assumed to be drawn independently from the same distribution $\mathrm{p}_{S,X}$ on $\mathcal{S} \times \mathcal{X}$ that is known to the aggregator through observing $(\vec{S}, \vec{X})$ (if one allows for non-independent X_i, then differential privacy is no longer an adequate privacy metric [15][8]). The aggregator, who has access to $\vec{X}$, sanitises the database by applying a sanitisation protocol (i.e., a random function) $\mathcal{Q}\colon \mathcal{X} \to \mathcal{Y}$ to each X_i, outputting $\vec{Y} = (Y_1, \dots, Y_n) = (\mathcal{Q}(X_1), \dots, \mathcal{Q}(X_n))$. The aggregator's goal is to find a $\mathcal{Q}$ that maximises the information about X_i preserved in Y_i (measured as $\mathrm{I}(X_i; Y_i)$) while leaking only minimal information about S_i.

Without loss of generality $\mathcal{X}, \mathcal{Y}, \mathcal{S}$ are identified with the sets $\{1, \dots, a\}, \{1, \dots, b\}, \{1, \dots, c\}$, respectively, for integers a, b, c. The subscript i from X_i, Y_i, S_i is omitted as no probabilities depend on it, and probabilities are written as p_x, p_s, $\mathrm{p}_{x|s}$, etc., which form vectors p_X, $\mathrm{p}_{S|x}$, etc., and matrices $\mathrm{p}_{X|S}$, etc.

As noted before, instead of looking at the mutual information $\mathrm{I}(S; Y)$, two different, related measures of sensitive information leakage known from the literature are considered. The first one is an adaptation of LDP, the *de facto* standard in information privacy [6]:

Definition 1. (ε-LDP) *Let $\varepsilon \in \mathbb{R}_{\geq 0}$. say that $\mathcal{Q}$ satisfies ε-LDP w.r.t. S if*

$$\forall y \in \mathcal{Y}, \forall s, s' \in \mathcal{S}: \quad \frac{\mathbb{P}(Y = y | S = s)}{\mathbb{P}(Y = y | S = s')} \leq \mathrm{e}^{\varepsilon}. \tag{1}$$

Most literature on LDP considers LDP w.r.t. X, i.e., for all y, x, x' it holds that

$$\frac{\mathbb{P}(Y = y | X = x)}{\mathbb{P}(Y = y | X = x')} \leq \mathrm{e}^{\varepsilon}. \tag{2}$$

This is a stricter requirement, because under this definition all data needs to be protected, rather than just the underlying sensitive data. This typically comes at a cost in utility [10].

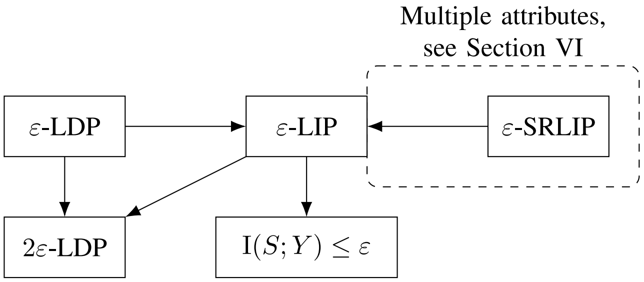


Figure 2. Relations between privacy notions. The multiple attributes setting is discussed in Section VI.

Throughout the present paper, ε-LDP always means ε-LDP w.r.t. S, unless otherwise specified.

The LDP metric reflects the fact that in the PF scenario one is only interested in hiding sensitive data, rather than all data; it is a specific case of what has been named Pufferfish Privacy [20]. The advantage of LDP compared to mutual information is that it gives privacy guarantees for the worst case, not just the average case. This is desirable in the database setting, as a worst-case metric guarantees the security of the private data of all users, while average-case metrics are only concerned with the average user. Another useful privacy metric is *Local Information Privacy* (LIP) [7][8], also called Removal Local Differential Privacy [19]:

Definition 2. (ε-LIP) *Let $\varepsilon \in \mathbb{R}_{\geq 0}$. The protocol $\mathcal{Q}$ satisfies ε-LIP w.r.t. S if*

$$\forall y \in \mathcal{Y}, s \in \mathcal{S}: \quad \mathrm{e}^{-\varepsilon} \leq \frac{\mathbb{P}(Y = y | S = s)}{\mathbb{P}(Y = y)} \leq \mathrm{e}^{\varepsilon}. \tag{3}$$

Compared to LDP, the disadvantage of LIP is that it depends on the distribution of S; this is not a problem in the PF scenario, as the aggregator, who chooses $\mathcal{Q}$, has access to the distribution of S. The advantage of LIP is that is more closely related to an attacker's capabilities: since

$$\frac{\mathbb{P}(Y = y | S = s)}{\mathbb{P}(Y = y)} = \frac{\mathbb{P}(S = s | Y = y)}{\mathbb{P}(S = s)}, \tag{4}$$

satisfying ε-LIP means that an attacker's posterior distribution of S given $Y = y$ does not deviate from their prior distribution by more than a factor e^{ε}. The following lemma outlines the relations between LDP, LIP and mutual information (see Figure 2).

Lemma 1. (See [8]) *Let $\mathcal{Q}$ be a sanitisation protocol, and let $\varepsilon \in \mathbb{R}_{\geq 0}$.*

1) *If $\mathcal{Q}$ satisfies ε-LDP, then it satisfies ε-LIP.*
2) *If $\mathcal{Q}$ satisfies ε-LIP, then it satisfies 2ε-LDP, and $\mathrm{I}(S; Y) \leq \varepsilon$.*

Remark 1. *One gets robust equivalents of LDP and LIP by demanding that $\mathcal{Q}$ satisfy ε-LIP (ε-LDP) for a set of distributions $\mathrm{p}_{S,X}$, instead of only a single distribution [20]. Letting $\mathrm{p}_{S,X}$ range over all possible distributions on $\mathcal{S} \times \mathcal{X}$ yields LIP (LDP) w.r.t. X.*

In this notation, instead of Problem 1 the following problem is considered:

Problem 2. *Suppose $\mathrm{p}_{S,X}$ is known to the aggregator, and let $\varepsilon \in \mathbb{R}_{\geq 0}$. Then, find the sanitisation protocol $\mathcal{Q}$ such that $\mathrm{I}(X;Y)$ is maximised while $\mathcal{Q}$ satisfies ε-LDP (ε-LIP, respectively) with respect to S.*

Note that this problem does not depend on the number of users n, and as such this approach will find solutions that are scalable w.r.t. n.

IV. Optimizing $\mathcal{Q}$ for ε-LDP

The goal is now to find the optimal $\mathcal{Q}$, i.e., the protocol that maximises $\mathrm{I}(X;Y)$ while satisfying ε-LDP, for a given ε. Any sanitisation protocol can be represented as a matrix $Q \in \mathbb{R}^{b\times a}$, where $Q_{y|x} = \mathbb{P}(Y = y|X = x)$. Then, ε-LDP is satisfied if and only if

$$\forall x\colon\ \sum_y Q_{y|x} = 1, \tag{5}$$

$$\forall x, y\colon\ 0 \leq Q_{y|x}, \tag{6}$$

$$\forall s, s', y\colon\ (Q\,\mathrm{p}_{X|s})_y \leq \mathrm{e}^{\varepsilon}(Q\,\mathrm{p}_{X|s'})_y. \tag{7}$$

As such, for a given $\mathcal{Y}$, the set of ε-LDP-satisfying sanitisation protocols can be considered a closed, bounded, convex polytope Γ in $\mathbb{R}^{b\times a}$. This fact allows us to efficiently find optimal protocols.

Theorem 2. *Let $\varepsilon \in \mathbb{R}_{\geq 0}$. Let $\mathcal{Q}\colon \mathcal{X} \to \mathcal{Y}$ be a ε-LDP protocol that maximises $\mathrm{I}(X;Y)$, i.e., the protocol that solves Problem 2 w.r.t. LDP.*

1) *One can take $b = a$.*
2) *Let Γ be the polytope described above, for $b = a$. Then the optimal $\mathcal{Q}$ corresponds to one of the vertices of Γ.*

Proof. The first result is obtained by generalising the results of [9]: there this is proven for regular ε-LDP (i.e., w.r.t. X), but the arguments given in that proof hold just as well in this situation; the only difference is that their polytope is defined by the ε-LDP conditions w.r.t. X, but this has no impact on the proof. The second statement follows from the fact that $\mathrm{I}(X;Y)$ is a convex function in $\mathcal{Q}$; therefore, its maximum on a bounded polytope is attained in one of the vertices. □

This theorem reduces the search for the optimal LDP protocol to enumerating the set of vertices of Γ, a $a(a-1)$-dimensional convex polytope. Note that the only property of $\mathrm{I}(X;Y)$ used in the proof is the fact that it is convex in Q. Therefore, the theorem holds for any convex utility metric.

One might argue that, since the optimal $\mathcal{Q}$ depends on $\mathrm{p}_{S,X}$, the publication of $\mathcal{Q}$ might provide an aggregator with information about the distribution of S. However, information on the distribution (as opposed to information of individual users' data) is not considered sensitive [23]. In fact, the reason why the aggregator sanitises the data is because an attacker is assumed to have knowledge about this correlation, and revealing too much information about X would cause the aggregator to use this information to infer information about S.

V. Optimizing $\mathcal{Q}$ for ε-LIP

If one uses ε-LIP as a privacy metric, one can find the optimal sanitisation protocol in a similar fashion. To do this, a sanitisation protocol $\mathcal{Q}$ is again described as a matrix, but this time a different one. Let $q \in \mathbb{R}^b$ be the probability mass function of Y, and let $R \in \mathbb{R}^{a\times b}$ be given by

$$R_{x|y} = \mathbb{P}(X = x|Y = y); \tag{8}$$

its y-th row is denoted by $R_{X|y} \in \mathbb{R}^a$. Then, a pair (R, q) defines a sanitisation protocol $\mathcal{Q}$ satisfying ε-LIP if and only if

$$\forall y\colon\ 0 \leq q_y, \tag{9}$$

$$Rq = \mathrm{p}_X, \tag{10}$$

$$\forall y\colon\ \sum_x R_{x|y} = 1, \tag{11}$$

$$\forall x, y\colon\ 0 \leq R_{x|y}, \tag{12}$$

$$\forall y, s\colon\ \mathrm{e}^{-\varepsilon}\,\mathrm{p}_s \leq \mathrm{p}_{s|X}\,R_{X|y} \leq \mathrm{e}^{\varepsilon}\,\mathrm{p}_s\,. \tag{13}$$

Note that (13) defines the ε-LIP condition, since for a given s, y one has

$$\frac{\mathrm{p}_{s|X}\,R_{X|y}}{\mathrm{p}_S} = \frac{\mathbb{P}(S = s|Y = y)}{\mathbb{P}(S = s)} = \frac{\mathbb{P}(Y = y|S = s)}{\mathbb{P}(Y = y)}. \tag{14}$$

(In)equalities (11–13) can be expressed as saying that for every $y \in \mathcal{Y}$ one has that $R_{X|y} \in \Delta$, where Δ is the convex closed bounded polytope in $\mathbb{R}^{\mathcal{X}}$ given by

$$\Delta = \left\{ v \in \mathbb{R}^{\mathcal{X}} :\ \begin{array}{l} \sum_x v_x = 1, \\ \forall x : 0 \leq v_x, \\ \forall s : \mathrm{e}^{-\varepsilon}\,\mathrm{p}_s \leq \mathrm{p}_{s|X}\,v \leq \mathrm{e}^{\varepsilon}\,\mathrm{p}_s \end{array} \right\}. \tag{15}$$

As in Theorem 2, this polytope can be used to find optimal protocols:

Theorem 3. *Let $\varepsilon \in \mathbb{R}_{\geq 0}$, and let Δ be the polytope above. Let $\mathcal{V} = \{v_1, \ldots, v_M\}$ be its set of vertices. For $v_i \in \mathcal{V}$, let $\mathrm{H}(v_i)$ be its entropy, i.e.*

$$\mathrm{H}(v_i) = -\sum_{x\in\mathcal{X}} v_{i,x} \ln(v_{i,x}). \tag{16}$$

Let $\hat{\alpha}$ be the solution to the optimisation problem

$$\begin{aligned} \text{minimise}_{\alpha\in\mathbb{R}^M}\quad & \sum_{i=1}^{M} \mathrm{H}(v_i)\alpha_i \\ \text{subject to}\quad & \forall i : \alpha_i \geq 0, \\ & \sum_{i=1}^{M} \alpha_i v_i = \mathrm{p}_X\,. \end{aligned} \tag{17}$$

Then the ε-LIP protocol $\mathcal{Q}\colon \mathcal{X} \to \mathcal{Y}$ that maximises $\mathrm{I}(X;Y)$ is given by

$$\mathcal{Y} = \{i \leq M : \hat{\alpha}_i > 0\}, \tag{18}$$

$$q_i = \hat{\alpha}_i, \tag{19}$$

$$R_{x|i} = v_{i,x}, \tag{20}$$

for all $i \in \mathcal{Y} \subseteq \{1, \ldots, M\}$ and all $x \in \mathcal{X}$. One has $b \leq a$.

Proof. This was proven for $\varepsilon = 0$ (i.e., when S and Y are independent) in [10], but the proof works similarly for $\varepsilon > 0$; the main difference is that the equality constraints of their (10) will be replaced by the inequality constraints of this paper's (13), but this has no impact on the proof presented there. □

Since linear optimisation problems can be solved fast, again the optimisation problem reduces to finding the vertices of a polytope. The advantage of using LIP instead of LDP is that Δ is a $(a-1)$-dimensional polytope, while Γ of Section IV is $a(a-1)$-dimensional. The time complexity of vertex enumeration is linear in the number of vertices [22], while the number of vertices can grow exponentially in the dimension of the polyhedron [11]. Together, this means that the dimension plays a huge role in the time complexity, hence the optimum under LIP is expected to be found significantly faster than under LDP.

VI. Multiple Attributes

An often-occuring scenario is that a user's data consists of multiple attributes, i.e.,

$$X = (X^1, \ldots, X^m) \in \mathcal{X} = \mathcal{X}^1 \times \cdots \times \mathcal{X}^m. \tag{21}$$

This can be problematic for this paper's approach for two reasons:

1) Such a large $\mathcal{X}$ can be problematic, since the computing time for optimisation both under LDP and LIP will depend heavily on a.
2) In practice, an attacker might sometimes utilise side channels to access some subsets of attributes X_i^j for some users. For these users, a sanitisation protocol can leak more information (w.r.t. to the attacker's updated prior information) than its LDP/LIP parameter would suggest.

To see how the second problem might arise in practice, suppose that X_i^1 is the height of individual i, X_i^2 is their weight, and S_i is whether i is obese or not. Since height is only lightly correlated with obesity, taking $Y_i = X_i^1$ would satisfy ε-LIP for some reasonably small ε. However, suppose that an attacker has access to X_i^2 via a side channel. While knowing i's weight gives the attacker some, but not perfect knowledge about i's obesity, the combination of the weight from the side channel, and the height from the Y_i, allows the attacker to calculate i's BMI, giving much more information about i's obesity. Therefore, the given protocol gives much less privacy in the presence of this side channel.

To solve the second problem, a more stringent privacy notion called *Side-channel Resistant LIP* (SRLIP) is introduced, which ensures that no matter which attributes an attacker has access to, the protocol still satisfies ε-LIP with respect to the attacker's new prior distribution. One could similarly introduce SRLDP, and many results will still hold for this privacy measure; nevertheless, since it has been concluded that LIP is preferable to LDP, the focus is on SRLIP. For any subset $J \subseteq \{1, \ldots, m\}$, the notation $\mathcal{X}^J$ is used for the set $\prod_{j \in J} \mathcal{X}^j$, and its elements are written as x^J.

Definition 3. (ε-SRLIP). *Let $\varepsilon > 0$, and let $\mathcal{X} = \prod_{j=1}^m \mathcal{X}^j$. The protocol $\mathcal{Q}$ satisfies ε-SRLIP if for every $y \in \mathcal{Y}$, for every $s \in \mathcal{S}$, for every $J \subseteq \{1, \ldots, m\}$, and for every $x^J \in \mathcal{X}^J$ one has*

$$\mathrm{e}^{-\varepsilon} \leq \frac{\mathbb{P}(Y = y | S = s, X^J = x^J)}{\mathbb{P}(Y = y | X^J = x^J)} \leq \mathrm{e}^{\varepsilon}. \tag{22}$$

In terms of Remark 1, $\mathcal{Q}$ satisfies ε-SRLIP if and only if it satisfies ε-LIP w.r.t. $\mathrm{p}_{S,X|x^J}$ for all J and x^J. Taking $J = \varnothing$ gives us the regular definition of ε-LIP, proving the following Lemma:

Lemma 2. *Let $\varepsilon > 0$. If $\mathcal{Q}$ satisfies ε-SRLIP, then $\mathcal{Q}$ satisfies ε-LIP.*

While SRLIP is stricter than LIP itself, it has the advantage that even when an attacker has access to some data of a user, the sanitisation protocol still does not leak an unwanted amount of information beyond the knowledge the attacker has gained via the side channel. Another advantage is that, contrary to LIP itself, SRLIP satisfies an analogon of the concept of *privacy budget* [14]:

Theorem 4. *Let $\mathcal{X} = \prod_{j=1}^m \mathcal{X}^j$, and for every j, let $\mathcal{Q}^j \colon \mathcal{X}^j \to \mathcal{Y}^j$ be a sanitisation protocol. Let $\varepsilon^j \in \mathbb{R}_{\geq 0}$ for every j. Suppose that for every $j \leq m$, for every $J \subseteq \{1, \ldots, j-1, j+1, \ldots, m\}$, and every $x^J \in \mathcal{X}^J$, $\mathcal{Q}^j$ satisfies ε^j-LIP w.r.t. $\mathrm{p}_{S,X|x^J}$. Then $\prod_j \mathcal{Q}^j \colon \mathcal{X} \to \prod_j \mathcal{Y}^j$ satisfies $\sum_j \varepsilon^j$-SRLIP.*

The proof is presented in Appendix A. This theorem tells us that to find a ε-SRLIP protocol for $\mathcal{X}$, it suffices to find a sanitisation protocol for each $\mathcal{X}^j$ that is $\frac{\varepsilon}{m}$-LIP w.r.t. a number of prior distributions. Unfortunately, the method of finding an optimal ε-LIP protocol w.r.t. one prior $\mathrm{p}_{S,X}$ of Theorem 3 does not transfer to the multiple prior setting. This is because this method only finds one (R, q), while by (10) a different (R, q) is needed for each prior distribution. Therefore, an approach similar to the one in Theorem 2 is adopted. The matrix Q^j (given by $Q^j_{y^j|x^j} = \mathbb{P}(\mathcal{Q}^j(x^j) = y^j)$) corresponding to $\mathcal{Q}^j : \mathcal{X}^j \to \mathcal{Y}^j$ satisfies the criteria of Theorem 4 if and only if the following criteria are satisfied:

$$\forall x^j : \sum_{y^j} Q^j_{y^j|x^j} = 1, \tag{23}$$

$$\forall x^j, y^j : 0 \leq Q^j_{y^j|x^j}, \tag{24}$$

$$\forall J, x^J, s, y^j : \mathrm{e}^{-\varepsilon/m} (Q^j\, \mathrm{p}_{X^j|x^J})_{y^j} \leq (Q^j\, \mathrm{p}_{X^j|s,x^J})_{y^j}, \tag{25}$$

$$\forall J, x^J, s, y^j : (Q^j\, \mathrm{p}_{X^j|s,x^J})_{y^j} \leq \mathrm{e}^{\varepsilon/m} (Q^j\, \mathrm{p}_{X^j|x^J})_{y^j}. \tag{26}$$

Similar to Theorem 2, the optimal Q^j satisfying these conditions can be found by finding the vertices of the polytope defined by (23–26). In terms of time complexity, the comparison to finding the optimal ε-LIP protocol via Theorem 3 versus finding a ε-SRLIP protocol via Theorem 4 is not straightforward. The complexity of enumerating the vertices of a polytope is $\mathcal{O}(ndv)$, where n is the number of inequalities, d is the dimension, and v is the number of vertices [22]. For the

Δ of Theorem 3 one has $d = a-1$ and $n = a+2c$. By contrast, the polytope defined by (23–26) satisfies $d = a^j(a^j - 1)$ and $n = (a^j)^2 + 2c\prod_{j' \neq j}(a^{j'} + 1)$. Finding v for both these polytopes is difficult, but in general $v \leq \binom{n}{d}$. Since this grows exponentially in d, Theorem 4 is expected to be faster when the a^j are small compared to a, i.e., when m is large. This will be investigated experimentally in Section VIII.

VII. Explicit protocols

The methods of Sections IV and V allow us to find the optimal LDP and LIP protocols. The complexity depends heavily on a and c, and can become computationally infeasible for large a and c. For such datasets, one has to rely on predetermined privacy algorithms. Two approaches are introduced: as a benchmark, Section VII-A discusses how 'standard' LDP protocols can be applied to the PF situation, and a new method, Conditional Reporting, that is meant to address the shortcomings of standard LDP protocols, is introduced in Section VII-B. As in the previous section, the focus is on LIP, but much of the discussion carries over to LDP as well.

A. Standard LDP protocols

In the literature, there are many examples of protocols $\mathcal{Q}\colon \mathcal{X} \to \mathcal{Y}$, depending on a privacy parameter α, whose output satisfies α-LDP with respect to X; for an overview see [24]. Such a protocol automatically satisfies α-LDP, hence certainly α-LIP, with respect to S. However, because X is only indirectly correlated with Y, such a protocol's actual LIP value may be better. The privacy of such a protocol $\mathcal{Q}$ is found by

$$\mathrm{LIP}(\mathcal{Q}) = \max_{y \in \mathcal{Y}, s \in \mathcal{S}} \left| \ln \frac{\sum_x Q_{y|x}\, \mathrm{p}_{x|s}}{\sum_x Q_{y|x}\, \mathrm{p}_x} \right|; \tag{27}$$

then $\mathcal{Q}$ satisfies ε-LIP if and only if $\mathrm{LIP}(\mathcal{Q}) \leq \varepsilon$.

This paper considers two LDP protocols. The first one is Generalised Rapid Response (GRR) [25]. The key strength of GRR is that for large enough α it maximises $\mathrm{I}(X;Y)$ [9]. Given α, GRR is a privacy protocol $\mathrm{GRR}^\alpha\colon \mathcal{X} \to \mathcal{X}$ given by

$$\mathrm{GRR}^\alpha_{y|x} = \begin{cases} \frac{\mathrm{e}^\alpha}{\mathrm{e}^\alpha + a - 1}, & \text{if } x = y, \\ \frac{1}{\mathrm{e}^\alpha + a - 1}, & \text{if } x \neq y. \end{cases} \tag{28}$$

A direct calculation then shows that

$$\mathrm{LIP}(\mathrm{GRR}^\alpha) = \max_{x,s} \left| \ln \frac{1 + (\mathrm{e}^\alpha - 1)\, \mathrm{p}_{x|s}}{1 + (\mathrm{e}^\alpha - 1)\, \mathrm{p}_x} \right|. \tag{29}$$

For GRR to satisfy ε-LIP, the equation $\mathrm{LIP}(\mathrm{GRR}^\alpha) = \varepsilon$ needs to be solved for α. Since $\mathrm{LIP}(\mathrm{GRR}^\alpha)$ is increasing in α, this can be done fast computationally.

The second protocol that is relevant to this paper is Optimised Unary Encoding (OUE) [26]. This protocol is notable for being one of the protocols that has the least known variance in frequency estimation [26]. For a choice of α as privacy parameter, and an input x, the output of $\mathrm{OUE}^\alpha\colon \mathcal{X} \to 2^{\mathcal{X}}$ is a vector of independent Bernoulli variables $E_{x'}$ for $x' \in \mathcal{X}$, satisfying

$$\mathbb{P}(E_{x'} = 1) = \begin{cases} \frac{1}{2}, & \text{if } x' = x, \\ \frac{1}{\mathrm{e}^\alpha + 1}, & \text{if } x' \neq x. \end{cases} \tag{30}$$

In other words, If a $y \in 2^{\mathcal{X}}$ is identified with a subset of $\mathcal{X}$ (so $\#y$ denotes its cardinality), one gets

$$\mathrm{OUE}^\alpha_{y|x} = \begin{cases} \frac{\mathrm{e}^{(a - \#y)\alpha}}{2(\mathrm{e}^\alpha + 1)^{a-1}}, & \text{if } x \in y, \\ \frac{\mathrm{e}^{(a - \#y - 1)\alpha}}{2(\mathrm{e}^\alpha + 1)^{a-1}}, & \text{if } x \notin y. \end{cases} \tag{31}$$

It follows that

$$\mathrm{LIP}(\mathrm{OUE}^\alpha) = \max_{y,s} \left| \ln \frac{1 + (\mathrm{e}^\alpha - 1) \sum_{x \in y} \mathrm{p}_{x|s}}{1 + (\mathrm{e}^\alpha - 1) \sum_{x \in y} \mathrm{p}_x} \right|. \tag{32}$$

B. Conditional Reporting

In general, a generic LDP protocol will not be ideal for the PF scenario, since these are designed to obscure all information about X, rather than just the part that holds information about S. To address this shortcoming, the protocol *Conditional Reporting* (CR) is introduced in Algorithm 1. This mechanism needs both S and X as input; hence it differs from the other protocols discussed in this paper, which only have X as input. The value of S is masked by Randomised Response. If the output $\tilde{s}$ equals S, the algorithm returns the true value of X. If not, it outputs a random one, whose probability distribution is given by $\mathrm{p}_{X|\tilde{s}}$.

Algorithm 1: Conditional Reporting (CR^α)

```
Input  : Privacy parameter $\alpha$; Probability distribution
         $\mathrm{p}_{S,X}$; input $(s, x) \in \mathcal{S} \times \mathcal{X}$
Output: $y \in \mathcal{X}$

Sample $\tilde{s} \in \mathcal{S}$ with
  $\mathbb{P}(\tilde{s} = s') = \begin{cases} \frac{\mathrm{e}^\alpha}{\mathrm{e}^\alpha + \#\mathcal{S} - 1}, & \text{if } s' = s, \\ \frac{1}{\mathrm{e}^\alpha + \#\mathcal{S} - 1}, & \text{otherwise} \end{cases}$
if $\tilde{s} = s$ then
    $y \leftarrow x$;
else
    Sample $\tilde{x} \in \mathcal{X}$ with $\mathbb{P}(\tilde{x} = x') = \mathrm{p}_{x'|\tilde{s}}$;
    $y \leftarrow \tilde{x}$;
end
```

CR^α certainly satisfies α-LDP, hence α-LIP, w.r.t. S. However, if S and X are not perfectly correlated, better privacy can be achieved, as outlined by the proposition below.

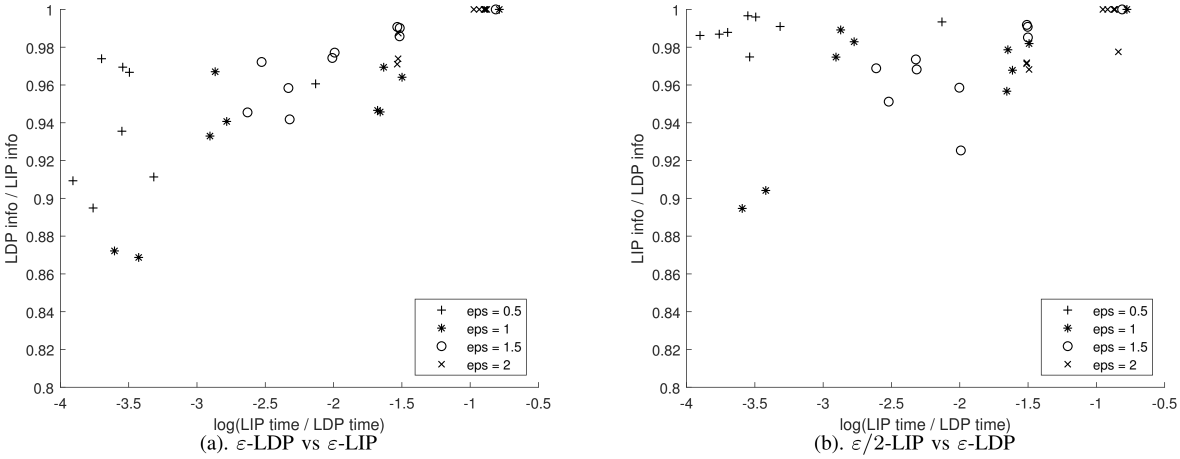
Proposition 1. *Given a probability distribution $\mathrm{p}_{X,S}$ and a $\alpha \geq 0$, define*

$$L(\alpha) = \max_{x,s} \left| \ln \frac{(\mathrm{e}^\alpha - 1)\, \mathrm{p}_{x|s} + \sum_{s'} \mathrm{p}_{x|s'}}{(\mathrm{e}^\alpha - 1)\, \mathrm{p}_x + \sum_{s'} \mathrm{p}_{x|s'}} \right|. \tag{33}$$

Then CR^α satisfies ε-LIP if and only if $\varepsilon \geq L(\alpha)$.

The proof is presented in Appendix A. One can use this proposition to find the α needed to have CR^α satisfy ε-LDP, by solving $L(\alpha) = \varepsilon$. At the very least one has the following upper bound:

Proposition 2. *The protocol CR^α satisfies α-LDP. In particular, it satisfies α-LIP, and $L(\alpha) \leq \alpha$.*



(a). ε-LDP vs ε-LIP

(b). $\varepsilon/2$-LIP vs ε-LDP

Figure 3. Comparison of computation time and $\mathrm{I}(X;Y)$ for LDP protocols found via Theorem 2 and LIP protocols found via Theorem 3, for random $\mathrm{p}_{S,X}$ with $c=2$, $a=5$, and $\varepsilon \in \{0.5, 1, 1.5, 2\}$.

Proof. For all $y \in \mathcal{X}$ and $s \in \mathcal{S}$ one has, following equation (48) in Appendix A, that

$$\mathbb{P}(\mathrm{CR}^{\alpha}(X,S)=y|S=s) = \tfrac{1}{\mathrm{e}^{\alpha}+c-1}\Big(\mathrm{e}^{\alpha}\,\mathrm{p}_{y|s} + \sum_{s' \neq s} \mathrm{p}_{y|s'}\Big). \tag{34}$$

It follows that

$$\begin{aligned} &\frac{\mathbb{P}(\mathrm{CR}^{\alpha}(X,S)=y|S=s)}{\mathbb{P}(\mathrm{CR}^{\alpha}(X,S)=y|S=s')} \\ &= \frac{\mathrm{e}^{\alpha}\,\mathrm{p}_{y|s} + p_{y|s'} + \sum_{s''\neq s,s'} \mathrm{p}_{y|s''}}{\mathrm{p}_{y|s} + \mathrm{e}^{\alpha} p_{y|s'} + \sum_{s''\neq s,s'} \mathrm{p}_{y|s''}} &(35)\\ &\leq \max\left\{1, \frac{\mathrm{e}^{\alpha}\,\mathrm{p}_{y|s} + p_{y|s'}}{\mathrm{p}_{y|s} + \mathrm{e}^{\alpha} p_{y|s'}}\right\} &(36)\\ &\leq \mathrm{e}^{\alpha}. \end{aligned}$$

□

VIII. Experiments

The feasibility of the different methods is tested by performing small-scale experiments on synthetic data and real-world data. All experiments are implemented in Matlab and conducted on a PC with Intel Core i7-7700HQ 2.8GHz and 32GB memory.

A. Synthetic data: LDP vs LIP

The computing time for finding optimal ε-LDP and ε-LIP protocols was compared for $c=2$ and $a=5$ for 10 random distributions $\mathrm{p}_{S,X}$, obtained by generating each $\mathrm{p}_{s,x}$ uniformly from $[0,1]$ and then normalising. The LDP/LIP privacy parameter ε is taken to be in $\{0.5, 1, 1.5, 2\}$; the results are in Figure 3(a). As one can see, Theorem 3 gives significantly faster results than Theorem 2; the average computing time for Theorem 2 for $\varepsilon = 0.5$ is 133s, while for Theorem 3 this is 0.0206s. With regards to the utility $\mathrm{I}(X;Y)$, since ε-LDP implies ε-LIP, the optimal ε-LIP protocol will have better utility than the optimal ε-LDP protocol. However, as can be seen from the figure, the difference in utility is relatively low.

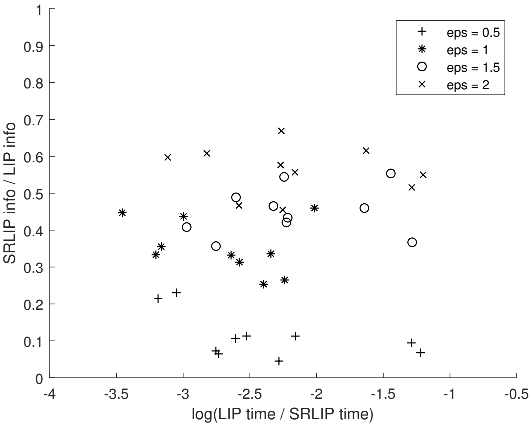


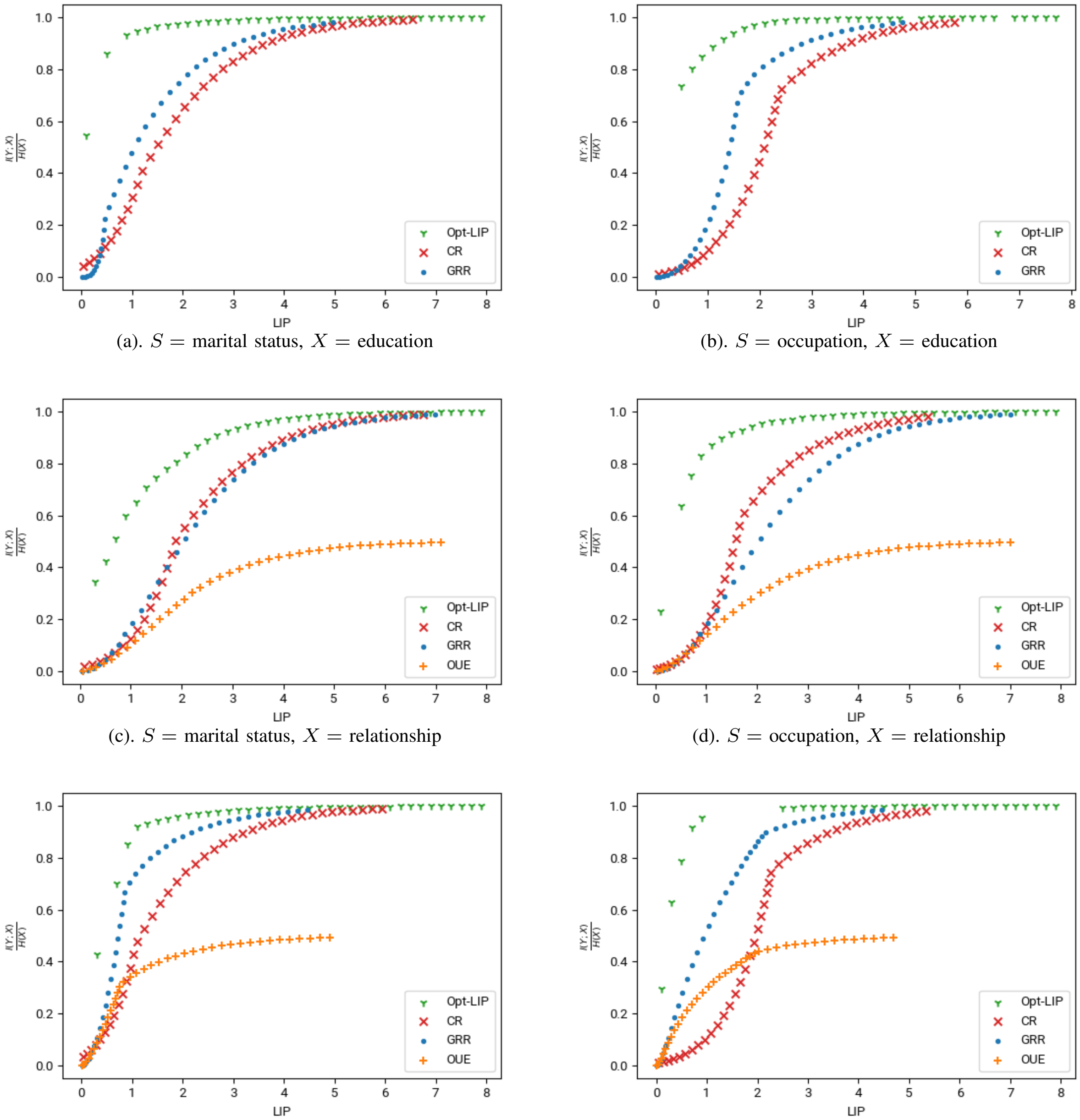
Figure 4. Comparison of computation time and $\mathrm{I}(X;Y)$ for ε-(SR)LIP-protocols found via Theorems 3 and 4, for random $\mathrm{p}_{S,X}$ with $c=2$, $a_1=a_2=3$, $a_3=4$, and $\varepsilon \in \{0.5, 1, 1.5, 2\}$.

Note that for bigger ε, both the difference in computing time and the difference in $\mathrm{I}(X;Y)$ between LDP and LIP become less. This is because of the probabilistic relation between S and X, for ε large enough, any sanitisation protocol satisfies ε-LIP and ε-LDP. This means that as ε grows, the resulting polytopes will have fewer defining inequalities, hence they will have fewer vertices. This results in lower computation times, which affects LDP more than LIP. At the same time, the fact that every protocol is both ε-LIP and ε-LDP will result in the same optimal utility.

In Figure 3(b), optimal $\frac{\varepsilon}{2}$-LDP protocols are compared to to optimal ε-LIP protocols. Again, LIP is significantly faster than LDP. Since ε-LIP implies $\frac{\varepsilon}{2}$-LDP, the optimal $\frac{\varepsilon}{2}$-LDP has higher utility; again the difference is low.

B. Synthetic data: LIP vs SRLIP

Similar comparisons are perfomed for multiple attributes, for $c=2$, $a_1=a_2=3$ and $a_3=4$, comparing the methods



(a). S = marital status, X = education

(b). S = occupation, X = education

(c). S = marital status, X = relationship

(d). S = occupation, X = relationship

(e). S = marital status, X = sex

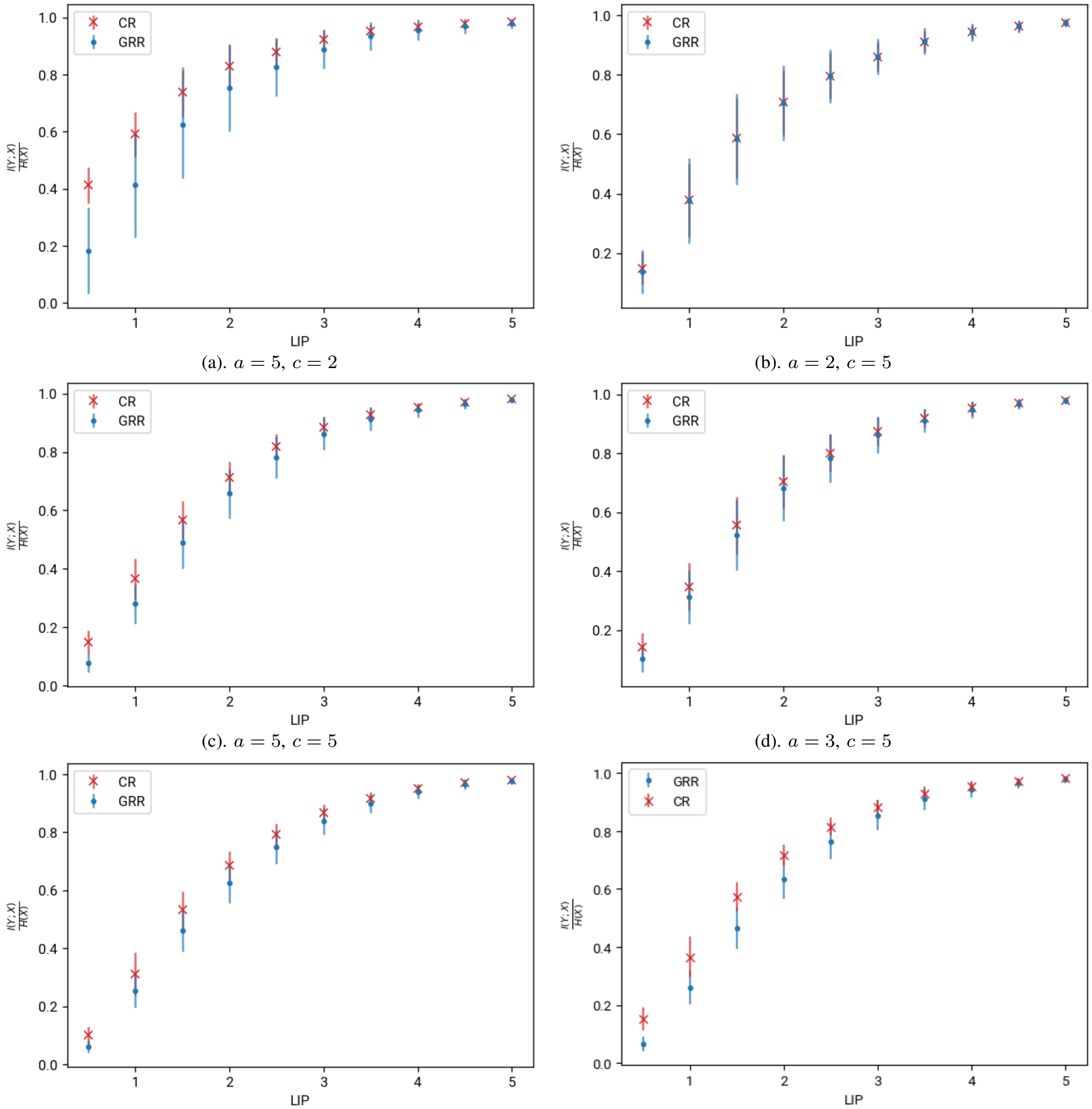
(f). S = occupation, X = sex

Figure 5. Experiments on the adult-dataset.

of Theorems 3 and 4. The results are presented in Figure 4. As one can see, Theorem 4 is significantly slower, with Theorem 3 being on average 476 times as fast. There is a sizable difference in utility, caused on one hand by the fact that ε-SRLIP is a stricter privacy requirement than ε-LIP, and on the other hand by the fact that Theorem 4 does not give us the optimal ε-SRLIP protocol.

C. Adult-dataset

The utility of Conditional Reporting (CR) is tested both on real world data and synthetic data. The real world data is from the well-known adult-dataset [27], which contains demographic data from the 1994 US census. For these experiments S is taken to be in $\{\text{marital status, occupation}\}$ (with $c = 7$ and $c = 15$, respectively) and X is taken to be in $\{\text{education, relationship, sex}\}$ (with $a = 16, 6, 2$). Based on the findings in the previous sections, LIP is taken as a privacy measure, and $\mathrm{I}(X;Y)$ as a utility measure. CR is compared on the one hand with the optimal method (Opt-LIP) found in Section V, and on the other hand with the established LDP protocols GRR and OUE. The results are shown in Figure 5. For X = education, the mutual information for OUE was infeasible to compute. Similarly, for S = occupation, some cases of Opt-LIP failed to compute within a reasonable



(a). $a = 5$, $c = 2$

(b). $a = 2$, $c = 5$

(c). $a = 5$, $c = 5$

(d). $a = 3$, $c = 5$

(e). $a = 5$, $c = 7$

(f). $a = 7$, $c = 5$

Figure 6. Experiments on synthetic data. For each value of a and c, the average utility is taken over 100 randomly generated probability distributions. Bar size denotes standard deviation.

timeframe. Nevertheless, it can concluded that GRR and CR both perform somewhere between Opt-LIP and OUE. As the LIP value ε grows larger, GRR and CR grow close to Opt-LIP. At the same time, OUE falls off for large ε, having $\frac{1}{2}\,\mathrm{H}(X)$ as its limit. This is because OUE by design only has probability $\frac{1}{2}$ transmitting the true X (as element of the set Y). The difference between GRR and CR is less clear, and it appears to depend on the joint distribution $\mathrm{p}_{X,S}$ which protocol gives the best utility.

D. Synthetic data: GRR vs CR

To investigate the difference between GRR and CR, both methods are applied to synthetic data. OUE is disregarded as it performs worse than the other two protocols, especially in the low privacy regime. For a fixed choice of a and c, 100 probability distributions are drawn from the Jeffreys prior on $\mathcal{S}\times\mathcal{X}$, i.e., the symmetric Dirichlet distribution with parameter $\frac{1}{2}$. A set of LIP values ε is fixed, and for each of these and each probability distribution, equations (29) and (33) are solved, setting the left hand side equal to ε and solving for α_{GRR} and α_{CR}. The mutual information $\mathrm{I}(X;Y)$ is then calculated, which is normalised by dividing by $\mathrm{H}(X)$. The resulting averages and standard deviations are displayed in Figure 6. On the whole, it can be seen that the larger a is compared to c, the more utility CR provides compared to GRR. However, this does not tell the whole story, as the difference between datasets has more impact on the utility than the difference between methods.

E. GRR and CR parameter α

To investigate what property of the probability distribution p_{XS} causes CR to outperform GRR, the parameters α_{CR} and α_{GRR} that govern the privacy protocols CR and GRR are considered. Both of these have the property that the higher their value, the less 'random' the protocols are, resulting in a better utility. Since these α are found from ε through different equations, the difference in utility of GRR and CR for different probability distributions may be explained by a difference in α. This assertion is tested for 100 randomly generated distributions in Figure 7. As can be seen, the difference in mutual information can for a large part be explained by a difference in α ($\rho = 0.9815$, $\rho = 0.9889$, and $\rho = 0.9731$, respectively). In Figure 8, the relation between α and the LIP value ε for the experiments in 5(b) and 5(d) is shown. The fact that $\alpha_{\mathrm{GRR}} > \alpha_{\mathrm{CR}}$ in 8(a) corresponds to the fact that GRR outperforms CR in 5(b), and the opposite relation holds between 8(b) and 5(d).

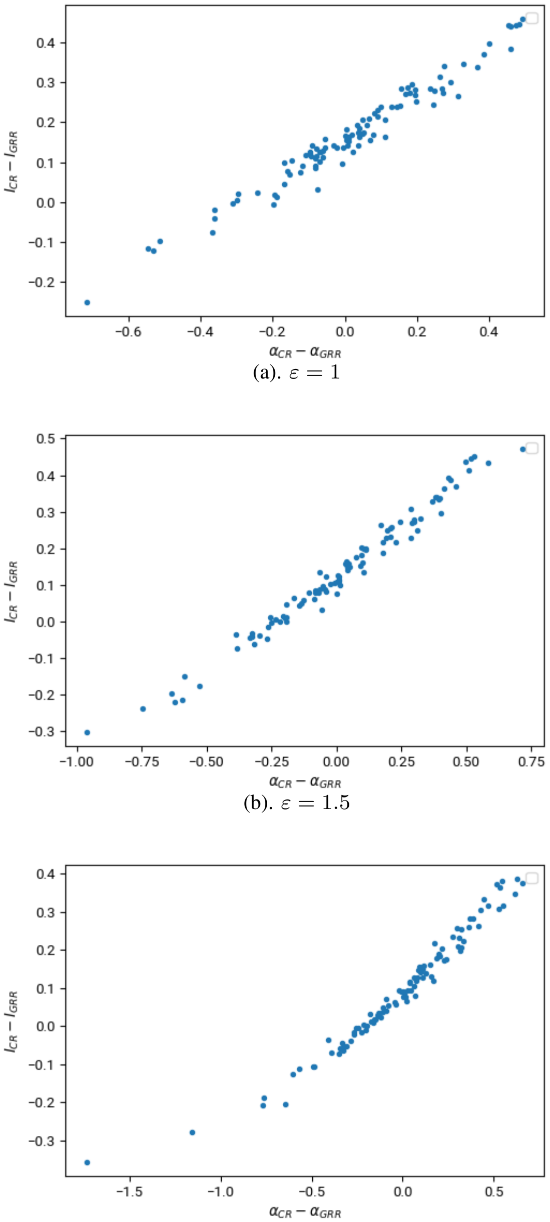
Unfortunately, we were not able to relate the difference in parameter α to other properties of the distribution. Without presenting details we mention that the properties $\mathrm{I}(X;S), \max_{x,s} \mathrm{p}_{x,s}, \max_x \mathrm{p}_x$ and $\max_s \mathrm{p}_s$ do not appear to have an impact on the difference in utility between GRR and CR.

IX. Conclusions and future work

Local data sanitisation protocols have the advantage of being scalable for large numbers of users. Furthermore, the advantage of using differential privacy-like privacy metrics is that they provide worst-case guarantees, ensuring that the privacy of every user is sufficiently protected. For both ε-LDP and ε-LIP methods are derived to find sanitisation protocols that maximise mutual information between input and output, solving the PF problem for these metrics.

Within this setting, it can be observed that ε-LIP has two main advantages over ε-LDP. First, it fits better within the PF setting, where the distribution $\mathrm{p}_{S,X}$ is (at least approximately) known to the estimator. Second, finding the optimal protocol is significantly faster than under LDP, especially for small ε. If one nevertheless prefers ε-LDP as a privacy metric, then it is still worthwile to find the optimal $\frac{\varepsilon}{2}$-LIP protocol, as this can be found significantly faster, at a low utility penalty.

In the multiple attributes setting, it is shown that ε-SRLIP provides additional privacy guarantees compared to ε-LIP, since without this requirement a protocol can lose all its privacy protection in the presence of side channels. Unfortunately, however, experiments show that this is paid for both in computation time and in utility.



(a). $\varepsilon = 1$

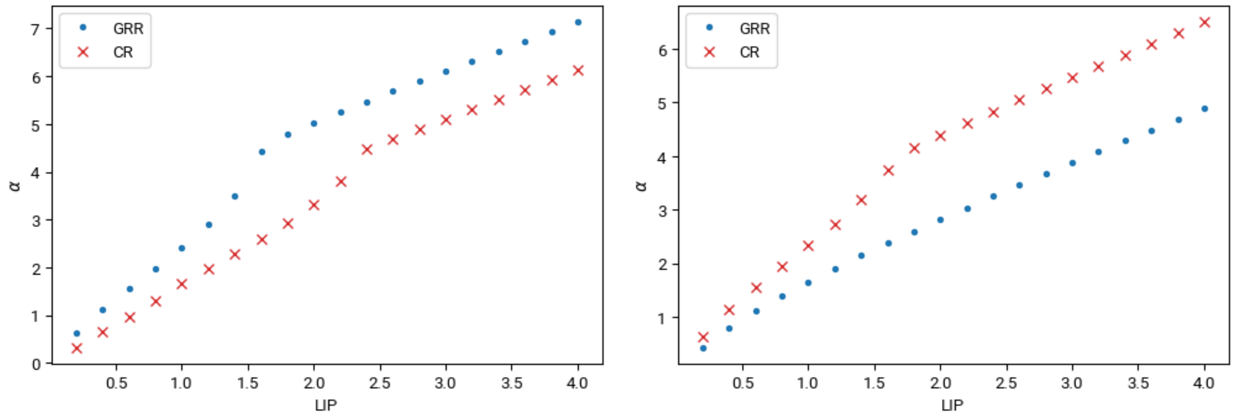
(b). $\varepsilon = 1.5$

(c). $\varepsilon = 2$

Figure 7. Difference in α versus difference in utility for 100 randomly generated probability distributions, for $a = c = 5$.

With regard to the specific protocols, it is found that the newly introduced protocol, CR, generally outperforms OUE, especially for high values of ε-LIP. This can be explained from the fact that by design the utility of OUE is capped at $\frac{1}{2}\,\mathrm{H}(X)$. CR behaves more or less similar to GRR, and which of these two protocols performs best depends on properties of the joint distribution $\mathrm{p}_{X,S}$. In particular, it largely depends on which of the two protocols has the highest value of their governing parameter α. Also, it can be seen that CR performs better on average if a is large compared to c.

For further research, a number of important avenues remain



(a) S = occupation, X = education

(b) S = occupation, X = relationship

Figure 8. Value of GRR and CR parameter α for different values of ε for the adult-dataset.

to be explored. First, the aggregator's knowledge about $\mathrm{p}_{S,X}$ may not be perfect, because they may learn about $\mathrm{p}_{S,X}$ through observing $(\vec{S},\vec{X})$. Incorporating this uncertainty leads to robust optimisation [28], which would give stronger privacy guarantees.

Second, it might be possible to improve the method of obtaining ε-SRLIP protocols via Theorem 4. Examining its proof shows that lower values of ε^j may suffice to still ensure ε-SRLIP. Furthermore, the optimal choice of $(\varepsilon^j)_{j\leq m}$ such that $\sum_j \varepsilon^j = \varepsilon$ might not be $\varepsilon^j = \frac{\varepsilon}{m}$. However, it is computationally prohibitive to perform the vertex enumeration for many different choices of $(\varepsilon^j)_{j\leq m}$, and as such a new theoretical approach is needed to determine the optimal $(\varepsilon^j)_{j\leq m}$ from ε and $\mathrm{p}_{S,X}$.

Third, it would be interesting to see if there are other ways to close the gap between the theoretically optimal protocol, which may be hard to compute in practice, and general LDP protocols, which do not see the difference between sensitive and non-sensitive information. This is relevant because CR needs both S and X as input, and there may be situations where access to S is not available.

Finally, although CR outperforms GRR and OUE for some datasets, it does not do so consistently. More research into the properties of distributions where CR fails to provide a significant advantage might lead to improved privacy protocols.

Acknowledgements

This work was supported by NWO grant 628.001.026 (Dutch Research Council, the Hague, the Netherlands).

References

[1] Milan Lopuhaä-Zwakenberg. "The Privacy Funnel from the Viewpoint of Local Differential Privacy". In: *Fourteenth International Conference on the Digital Society (ICDS)* (2020), pp. 19–24.

[2] Flavio du Pin Calmon, Ali Makhdoumi, Muriel Médard, Mayank Varia, Mark Christiansen, and Ken R Duffy. "Principal inertia components and applications". In: *IEEE Transactions on Information Theory* 63.8 (2017), pp. 5011–5038.

[3] Ni Ding and Parastoo Sadeghi. "A Submodularity-based Agglomerative Clustering Algorithm for the Privacy Funnel". In: *arXiv:1901.06629* (2019). Preprint, accessed 2020.11.8.

[4] Fabian Prasser, Florian Kohlmayer, Ronald Lautenschlaeger, and Klaus A. Kuhn. "Arx-a comprehensive tool for anonymizing biomedical data". In: *AMIA Annual Symposium Proceedings*. Vol. 2014. American Medical Informatics Association. 2014, p. 984.

[5] Ali Makhdoumi, Salman Salamatian, Nadia Fawaz, and Muriel Médard. "From the information bottleneck to the privacy funnel". In: *2014 IEEE Information Theory Workshop (ITW 2014)*. IEEE. 2014, pp. 501–505.

[6] Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. "What can we learn privately?" In: *SIAM Journal on Computing* 40.3 (2011), pp. 793–826.

[7] Bo Jiang, Ming Li, and Ravi Tandon. "Local Information Privacy with Bounded Prior". In: *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE. 2019, pp. 1–7.

[8] Salman Salamatian, Flavio du Pin Calmon, Nadia Fawaz, Ali Makhdoumi, and Muriel Médard. "Privacy-Utility Tradeoff and Privacy Funnel". In: *http://www.mit.edu/~salmansa/files/privacy_TIFS.pdf* (2020). Preprint, accessed 2020.11.8.

[9] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. "Extremal mechanisms for local differential privacy". In: *Advances in neural information processing systems*. 2014, pp. 2879–2887.

[10] Borzoo Rassouli and Deniz Gunduz. "On perfect privacy". In: *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2018, pp. 2551–2555.

[11] Imre Bárány and Attila Pór. "On 0-1 polytopes with many facets". In: *Advances in Mathematics* 161.2 (2001), pp. 209–228.

[12] Naftali Tishby, Fernando C Pereira, and William Bialek. "The information bottleneck method". In: *arXiv:physics/0004057* (2000). Preprint.

[13] SY Kung. "A compressive privacy approach to generalized information bottleneck and privacy funnel problems". In: *Journal of the Franklin Institute* 355.4 (2018), pp. 1846–1872.

[14] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. "Calibrating noise to sensitivity in private data analysis". In: *Theory of cryptography conference*. Springer. 2006, pp. 265–284.

[15] Paul Cuff and Lanqing Yu. "Differential privacy as a mutual information constraint". In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. 2016, pp. 43–54.

[16] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. "Our data, ourselves: Privacy via distributed noise generation". In: *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer. 2006, pp. 486–503.

[17] Cynthia Dwork and Guy N Rothblum. "Concentrated differential privacy". In: *arXiv:1603.01887* (2016). Preprint, accessed 2020.11.8.

[18] Ilya Mironov. "Rényi differential privacy". In: *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*. IEEE. 2017, pp. 263–275.

[19] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Shuang Song, Kunal Talwar, and Abhradeep Thakurta. "Encode, shuffle, analyze privacy revisited: formalizations and empirical evaluation". In: *arXiv:2001.03618* (2020). Preprint, accessed 2020.11.8.

[20] Daniel Kifer and Ashwin Machanavajjhala. "Pufferfish: A framework for mathematical privacy definitions". In: *ACM Transactions on Database Systems (TODS)* 39.1 (2014), pp. 1–36.

[21] Naoise Holohan, Douglas J Leith, and Oliver Mason. "Extreme points of the local differential privacy polytope". In: *Linear Algebra and its Applications* 534 (2017), pp. 78–96.

[22] David Avis and Komei Fukuda. "A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra". In: *Discrete & Computational Geometry* 8.3 (1992), pp. 295–313.

[23] Milan Lopuhaä-Zwakenberg, Boris Škorić, and Ninghui Li. "Information-theoretic metrics for Local Differential Privacy protocols". In: *arXiv:1910.07826* (2019). Preprint, accessed 2020.11.8.

[24] Mengmeng Yang, Lingjuan Lyu, Jun Zhao, Tianqing Zhu, and Kwok-Yan Lam. "Local Differential Privacy and Its Applications: A Comprehensive Survey". In: *arXiv:2008.03686* (2020). Preprint, 2020.11.8.

[25] Stanley L Warner. "Randomized response: A survey technique for eliminating evasive answer bias". In: *Journal of the American Statistical Association* 60.309 (1965), pp. 63–69.

[26] Tianhao Wang, Jeremiah Blocki, Ninghui Li, and Somesh Jha. "Locally differentially private protocols for frequency estimation". In: *26th {USENIX} Security Symposium ({USENIX} Security 17)*. 2017, pp. 729–745.

[27] Dheeru Dua and Casey Graff. *UCI Machine Learning Repository*. 2017. URL: http://archive.ics.uci.edu/ml/datasets/Adult.

[28] Dimitris Bertsimas, Vishal Gupta, and Nathan Kallus. "Data-driven robust optimization". In: *Mathematical Programming* 167.2 (2018), pp. 235–292.

Appendix A
Proofs

Proof of Theorem 4. For $J \subseteq \{1,\ldots,m\}$ and $j \in \{1,\ldots,m\}$, define $J[j] := J \cap \{1,\ldots,j-1\}$. Furthermore, write $\mathcal{X}^{\setminus J} = \prod_{j \notin J} \mathcal{X}^j$, and its elements as $x^{\setminus J}$. Define $\varepsilon := \sum_j \varepsilon^j$. Then

$$\begin{aligned}
\mathrm{p}_{y|s,x^J} &= \sum_{x^{\setminus J}} \mathrm{p}_{y|x}\, \mathrm{p}_{x^{\setminus J}|s,x^J} && (37)\\
&= \mathrm{p}_{y^J|x^J} \sum_{x^{\setminus j}} \left(\prod_{j \notin J} \mathrm{p}_{y^j|x^j} \right) \mathrm{p}_{x^{\setminus J}|s,x^J} && (38)\\
&= \mathrm{p}_{y^J|x^J} \sum_{x^{\setminus j}} \prod_{j \notin J} \mathrm{p}_{y^j|x^j}\, \mathrm{p}_{x^j|s,x^{J[j]}} && (39)\\
&= \mathrm{p}_{y^J|x^J} \prod_{j \notin J} \sum_{x^j} \mathrm{p}_{y^j|x^j}\, \mathrm{p}_{x^j|s,x^{J[j]}} && (40)\\
&= \mathrm{p}_{y^J|x^J} \prod_{j \notin J} \mathrm{p}_{y^j|s,x^{J[j]}} && (41)\\
&\leq \mathrm{p}_{y^J|x^J} \prod_{j \notin J} \mathrm{e}^{\varepsilon^j}\, \mathrm{p}_{y^j|x^{J[j]}} && (42)\\
&\leq \mathrm{e}^{\varepsilon}\, \mathrm{p}_{y^J|x^J} \prod_{j \notin J} \mathrm{p}_{y^j|x^{J[j]}} && (43)\\
&= \mathrm{e}^{\varepsilon}\, \mathrm{p}_{y|x^J}\,. && (44)
\end{aligned}$$

The fact that $\mathrm{e}^{-\varepsilon}\, \mathrm{p}_{y|x^J} \leq \mathrm{p}_{y|s,x^J}$ is proven analogously. □

Proof of Proposition 1. Write $Q_{y|x,s} = \mathbb{P}(\mathrm{CR}^\alpha(x,s) = y)$. Then

$$Q_{y|x,s} = \sum_{s'} \mathbb{P}(\mathrm{CR}^\alpha(x,s) = y|\tilde{s} = s')\mathbb{P}(\tilde{s} = s'|S = s) \quad (45)$$

$$= \frac{\mathrm{e}^\alpha}{\mathrm{e}^\alpha + c - 1} + \frac{1}{\mathrm{e}^\alpha + c - 1} \sum_{s' \neq s} \mathrm{p}_{y|s'}, \quad (46)$$

where $\delta_{x=y}$ is the Kronecker delta. It follows that

$$\begin{aligned}
&\mathbb{P}(\mathrm{CR}^\alpha(X,S) = y|S = s)\\
&= \sum_x Q_{y|x,s}\, \mathrm{p}_{x|s} && (47)
\end{aligned}$$

$$= \frac{\mathrm{e}^{\alpha}}{\mathrm{e}^{\alpha}+c-1}\,\mathrm{p}_{y|s} + \frac{1}{\mathrm{e}^{\alpha}+c-1}\sum_{s'\neq s}\mathrm{p}_{y|s'} \tag{48}$$

$$= \frac{\mathrm{e}^{\alpha}-1}{\mathrm{e}^{\alpha}+c-1}\,\mathrm{p}_{y|s} + \frac{1}{\mathrm{e}^{\alpha}+c-1}\sum_{s'}\mathrm{p}_{y|s'}, \tag{49}$$

$$\mathbb{P}(\mathrm{CR}^{\alpha}(X,S)=y)$$

$$= \sum_{s}\mathbb{P}(\mathrm{CR}^{\alpha}(X,S)=y|S=s)\,\mathrm{p}_{s} \tag{50}$$

$$= \frac{\mathrm{e}^{\alpha}}{\mathrm{e}^{\alpha}+c-1}\,\mathrm{p}_{y} + \frac{1}{\mathrm{e}^{\alpha}+c-1}\sum_{s}\sum_{s'\neq s}\mathrm{p}_{y|s'}\,\mathrm{p}_{s} \tag{51}$$

$$= \frac{\mathrm{e}^{\alpha}}{\mathrm{e}^{\alpha}+c-1}\,\mathrm{p}_{y} + \frac{1}{\mathrm{e}^{\alpha}+c-1}\sum_{s'}\mathrm{p}_{y|s'}\sum_{s\neq s'}\mathrm{p}_{s} \tag{52}$$

$$= \frac{\mathrm{e}^{\alpha}}{\mathrm{e}^{\alpha}+c-1}\,\mathrm{p}_{y} + \frac{1}{\mathrm{e}^{\alpha}+c-1}\sum_{s'}(\mathrm{p}_{y|s'}-\mathrm{p}_{y,s'}) \tag{53}$$

$$= \frac{\mathrm{e}^{\alpha}-1}{\mathrm{e}^{\alpha}+c-1}\,\mathrm{p}_{y} + \frac{1}{\mathrm{e}^{\alpha}+c-1}\sum_{s'}\mathrm{p}_{y|s'}\,. \tag{54}$$

It follows that

$$L(\alpha) = \max_{y,s}\left|\ln\frac{\mathbb{P}(\mathrm{CR}^{\alpha}(X,S)=y|S=s)}{\mathbb{P}(\mathrm{CR}^{\alpha}(X,S)=y)}\right|, \tag{55}$$

hence CR^{α} satisfies ε-LIP if and only if $\varepsilon \geq L(\alpha)$. □