Data Sanitisation Protocols for the Privacy Funnel with Differential Privacy Guarantees

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. We tackle these problems by using differential privacy metrics, and by considering local protocols which operate on one entry at a time. We show that under both the Local Differential Privacy and Local Information Privacy leakage metrics, one can efficiently obtain optimal protocols. Furthermore, Local Information Privacy is both more closely aligned to the privacy requirements of the Privacy Funnel scenario, and more efficiently computable. We also consider the scenario where each user has multiple attributes, for which we define *Side-channel Resistant Local Information Privacy*, and we give efficient methods to find protocols satisfying this criterion while still offering good utility. Finally, we introduce *Conditional Reporting*, an explicit LIP protocol that can be used when the optimal protocol is infeasible to compute, and we test this protocol on real-world and synthetic data. Experiments on real-world and synthetic data confirm the validity of these methods.**

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

I. Introduction

This paper is an extended version of [18]. 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 $\vec{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$ 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 $X_i = (X_i^1, \ldots, X_i^m)$; we will consider this in more detail in Section V). 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 without leaking the S_i, the aggregator releases a privatised database $\vec{Y} = (Y_1, \ldots, Y_n)$, obtained from applying a sanitisation mechanism $\mathcal{R}$ to $\vec{X}$. One way to formulate this is by considering the *Privacy Funnel*:

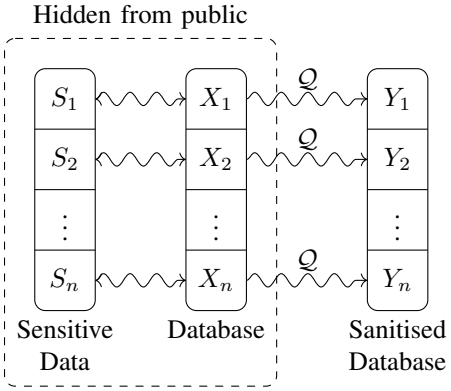


Figure 1. *Model of the Privacy Funnel with local protocols.*

Problem 1. (Privacy Funnel, [4]) *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 privatization mechanisms that operate on all of $\vec{X}$ can be computationally prohibitive for large n, as the complexity is exponential in n [6][22].
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, we make two changes to the general approach. First, we look at *local* data sanitisation, i.e., we consider optimization protocols $\mathcal{Q}\colon \mathcal{X} \to \mathcal{Y}$, for some finite set $\mathcal{Y}$, and we apply $\mathcal{Q}$ 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 Privacy Funnel setting [20][6]. Second, to ensure strong privacy guarantees even in worst-case scenarios, we take stricter notions of privacy, based on Local Differential Privacy (LDP) [15]. For these metrics, we develop methods

to find optimal protocols. Furthermore, for situations where the optimal protocol is computationally unfeasible to find, we introduce a new protocol, *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, we adapt two Differential Privacy-like privacy metrics to the Privacy Funnel situation, namely Local Differential Privacy (LDP) [15] and Local Information Privacy (LIP) [13][24]. We modify these metrics so that they measure leakage about the underlying S rather than X itself (for notational convenience, we write S, X, Y rather than S_i, X_i, Y_i throughout the rest of this paper). For a given level of leakage, we are interested in the privacy protocol that maximises the mutual information between input X_i and output Y_i. Adapting methods from [14] on LDP and [23] on perfect privacy, we prove the following Theorem:

Theorem 1 (Theorems 2 and 3 paraphrased)**.** *Suppose $a = \#\mathcal{X}, c = \#\mathcal{S}$, and $\mathrm{p}_{X,S}$, as well as 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.*

The descriptions of these polytopes, and how they relate to the optimisation problem, are discussed in Sections III and IV, respectively. Since the complexity of the polytope vertex enumeration depends significantly on both its dimension and the number of defining inequalities [2], finding optimal LIP protocols can be done significantly faster than finding optimal LDP protocols. Furthermore, we will argue that LIP is a privacy metric that more accurately captures information leakage than LDP in the Privacy Funnel scenario. For these two reasons we only consider LIP 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, we introduce a new privacy metric called *Sidechannel-Resistant LIP* that takes such sidechannels into account. We expand the vertex enumeration methods outlined above to find optimal SRLIP methods in Section V.

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 our 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, we introduce Conditional Reporting (CR), a privacy protocol that focuses on hiding S rather than X, in Section VI. Finding the appropriate CR protocol for a given probability distribution and privacy level can be done fast numerically.

In Section VII, we test the methods and protocols discussed above on both synthetic and real data. Compared to [18], new content in this extended paper are Section VI, the experiments on real data, and the extended literature review.

B. Related work

The Privacy Funnel (PF) setting was introduced in [20], 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. The Privacy Funnel is related to the Information Bottleneck (IB) [25], 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 [17][6]. A wider range of privacy metrics for the Privacy Funnel, and their relation to Differential Privacy, is discussed in [24].

Local Differential Privacy (LDP) was introduced in [15]. It is an adaptation of Differential Privacy [9] 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 (Local) Differential Privacy have been proposed [5][10][8][21]. We are particularly interested in Local Information Privacy (LIP) [13][24], also called Removal Local Differential Privacy [11]. 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 [16].

In [14], 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 [12]. Similarly, [23] 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 [1][2].

II. Mathematical Setting

The database $\vec{X} = (X_1, \ldots, X_n)$ 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 we assume $\mathcal{S}$ to be finite (see Figure 1). We assume that each (S_i, X_i) is drawn independently from

the same distribution $\mathrm{p}_{S,X}$ on $\mathcal{S} \times \mathcal{X}$ which 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 [5][24]). 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 $\vec{Y} = (Y_1, \ldots, Y_n) = (\mathcal{Q}(X_1), \ldots, \mathcal{Q}(X_n))$. 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 we write $\mathcal{X} = \{1, \ldots, a\}$, $\mathcal{Y} = \{1, \ldots, b\}$ and $\mathcal{S} = \{1, \ldots, c\}$ for integers a, b, c. We omit the subscript i from X_i, Y_i, S_i as no probabilities depend on it, and we write such probabilities 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 $\mathrm{I}(S;Y)$, we consider two different, related measures of sensitive information leakage known from the literature. The first one is an adaptation of LDP, the *de facto* standard in information privacy [15]:

Definition 1. (ε-LDP) *Let $\varepsilon \in \mathbb{R}_{\geq 0}$. We 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. $\frac{\mathbb{P}(Y=y|X=x)}{\mathbb{P}(Y=y|X=x')} \leq \mathrm{e}^{\varepsilon}$ for all x, x', y. Throughout this paper, by ε-LDP we always mean ε-LDP w.r.t. S, unless otherwise specified.

The LDP metric reflects the fact that we are only interested in hiding sensitive data, rather than all data; it is a specific case of what has been named 'pufferfish privacy' [16]. 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) [13][24], also called Removal Local Differential Privacy [11]:

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

$$\forall_{y \in \mathcal{Y}, s \in \mathcal{S}} \qquad \mathrm{e}^{-\varepsilon} \leq \frac{\mathbb{P}(Y = y | S = s)}{\mathbb{P}(Y = y)} \leq \mathrm{e}^{\varepsilon}. \tag{2}$$

Compared to LDP, the disadvantage of LIP is that it depends on the distribution of S; this is not a problem in our 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 $\frac{\mathbb{P}(Y=y|S=s)}{\mathbb{P}(Y=y)} = \frac{\mathbb{P}(S=s|Y=y)}{\mathbb{P}(S=s)}$, 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).

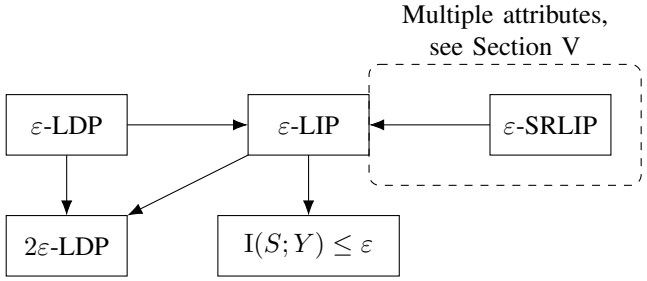


Figure 2. *Relations between privacy notions. The multiple attributes setting is discussed in Section V.*

Lemma 1. (See [24]) *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 [16]. 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 we consider the following problem:

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.

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

Our 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. We can represent any sanitisation protocol 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

$$\forall x\colon \ \sum_y Q_{y|x} = 1, \tag{3}$$

$$\forall x, y\colon \ 0 \leq Q_{y|x}, \tag{4}$$

$$\forall s, s', y\colon \ (Q\,\mathrm{p}_{X|s})_y \leq \mathrm{e}^{\varepsilon} (Q\,\mathrm{p}_{X|s'})_y. \tag{5}$$

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 [14]: 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 our 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.

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 [19]. 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.

IV. 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, we again describe $\mathcal{Q}$ 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 $R_{x|y} = \mathbb{P}(X = x|Y = y)$; we denote its y-th row 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

$$\forall y\colon\ 0 \le q_y, \tag{6}$$

$$Rq = \mathrm{p}_X, \tag{7}$$

$$\forall y\colon\ \sum_x R_{x|y} = 1, \tag{8}$$

$$\forall x, y\colon\ 0 \le R_{x|y}, \tag{9}$$

$$\forall y, s\colon\ \mathrm{e}^{-\varepsilon}\,\mathrm{p}_s \le \mathrm{p}_{s|X}\,R_{X|y} \le \mathrm{e}^{\varepsilon}\,\mathrm{p}_s\,. \tag{10}$$

Note that (10) defines the ε-LIP condition, since for a given s, y we have $\frac{\mathrm{p}_{s|X}\,R_{X|y}}{\mathrm{p}_S} = \frac{\mathbb{P}(S=s|Y=y)}{\mathbb{P}(S=s)} = \frac{\mathbb{P}(Y=y|S=s)}{\mathbb{P}(Y=y)}$. (In)equalities (8–10) 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

$$\Delta = \left\{ v \in \mathbb{R}^{\mathcal{X}} :\ \begin{array}{l} \sum_x v_x = 1, \\ \forall x : 0 \le v_x, \\ \forall s : \mathrm{e}^{-\varepsilon}\,\mathrm{p}_s \le \mathrm{p}_{s|X}\,v \le \mathrm{e}^{\varepsilon}\,\mathrm{p}_s \end{array} \right\}. \tag{11}$$

As in Theorem 2, we can use this polytope to find optimal protocols:

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

$$\mathrm{H}(v_i) = -\sum_{x\in\mathcal{X}} v_x \ln(v_{i,x}). \tag{12}$$

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

$$\text{minimise}_{\alpha\in\mathbb{R}^M} \quad \sum_{i=1}^{M} \mathrm{H}(v_i)\alpha_i \tag{13}$$

$$\text{subject to} \quad \forall i : \alpha_i \ge 0, \tag{14}$$

$$\sum_{i=1}^{M} \alpha_i v_i = \mathrm{p}_X\,. \tag{15}$$

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

$$\mathcal{Y} = \{i \le M : \hat{\alpha}_i > 0\}, \tag{16}$$

$$q_i = \hat{\alpha}_i, \tag{17}$$

$$R_{x|i} = v_{i,x}, \tag{18}$$

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

Proof. This was proven for $\varepsilon = 0$ (i.e., when S and Y are independent) in [23], 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 our (10), but this has no impact on the proof presented there. □

Since linear optimization problems can be solved fast, again the optimization 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 III is $a(a-1)$-dimensional. The time complexity of vertex enumeration is linear in the number of vertices [1], while the number of vertices can grow exponentially in the dimension of the polyhedron [2]. Together, this means that the dimension plays a huge role in the time complexity, hence we expect finding the optimum under LIP to be significantly faster than under LDP.

V. Multiple Attributes

An often-occuring scenario is that a user's data consists of multiple attributes, i.e., $X = (X^1, \dots, X^m) \in \mathcal{X} = \mathcal{X}^1 \times \cdots \times \mathcal{X}^m$. This can be problematic for our 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, we introduce a more stringent privacy notion called *Side-channel Resistant LIP* (SRLIP), 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 we concluded that LIP is preferable to LDP, we focus on SRLIP. For any subset $J \subseteq \{1, \dots, m\}$, we write $\mathcal{X}^J = \prod_{j \in J} \mathcal{X}^j$ and its elements as x^J.

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

$$\mathrm{e}^{-\varepsilon} \leq \frac{\mathbb{P}(Y = y | S = s, X^J = x^J)}{\mathbb{P}(Y = y | X^J = x^J)} \leq \mathrm{e}^{\varepsilon}. \tag{19}$$

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* [9]:

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 $J \subseteq \{1, \dots, j-1, j+1, \dots, m\}$, 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 (7) we need a different (R, q) for each prior distribution. Therefore, we are forced to adopt an approach similar to the one in Theorem 2. The matrix Q^j (given by $Q^j_{y^j|x^j} = \mathbb{P}(\mathcal{Q}^j(x^j) = y^j)$) corresponding to $\mathcal{Q}^j \colon \mathcal{X}^j \to \mathcal{Y}^j$ satisfies the criteria of Theorem 4 if and only if the following criteria are satisfied:

$$\forall x^j : \sum_{y^j} Q^j_{y^j|x^j} = 1, \tag{20}$$

$$\forall x^j, y^j : 0 \leq Q^j_{y^j|x^j}, \tag{21}$$

$$\forall J, x^J, s, y^j : \mathrm{e}^{-\varepsilon/m} (Q^j \, \mathrm{p}_{X^j|x^J})_{y^j} \leq (Q^j \, \mathrm{p}_{X^j|s,x^J})_{y^j}, \tag{22}$$

$$\forall J, x^J, s, y^j : (Q^j \, \mathrm{p}_{X^j|s,x^J})_{y^j} \leq \mathrm{e}^{\varepsilon/m} (Q^j \, \mathrm{p}_{X^j|x^J})_{y^j}. \tag{23}$$

Similar to Theorem 2, we can find the optimal $\mathcal{Q}^j$ satisfying these conditions by finding the vertices of the polytope defined by (20–23). 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 [1]. For the Δ of Theorem 3 we have $d = a - 1$ and $n = a + 2c$. In contrast, the polytope defined by (20–23) 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, we expect Theorem 4 to be faster when the a^j are small compared to a, i.e., when m is large. We will investigate this experimentally in the next section.

VI. Explicit protocols

The methods of Sections III and IV 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. We consider two approaches: as a benchmark, we discuss how 'standard' LDP protocols can be applied to the Privacy Funnel situation, and we introduce a new method, Conditional Reporting, that is meant to address the shortcomings of standard LDP protocols. As in the previous section, we focus 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 [28]. 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 lower. We can find the privacy of such a protocol $\mathcal{Q}$ by

$$\mathrm{LIP}(\mathcal{Q}) = \max_{y \in \mathcal{Y}, s \in \mathcal{S}} \left| \ln \frac{\sum_x Q_{y|x} \, \mathrm{p}_{x|s}}{\sum_x Q_{y|x} \, \mathrm{p}_x} \right|; \tag{24}$$

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

For this paper we are mainly interested in two protocols. The first one is Generalised Rapid Response (GRR) [27]. We are interested in GRR because for large enough α it maximises $\mathrm{I}(X;Y)$ [14]. Given α, GRR is a privacy protocol $\mathrm{GRR}^\alpha \colon \mathcal{X} \to \mathcal{X}$ given by

$$\mathrm{GRR}^\alpha_{y|x} = \begin{cases} \frac{\mathrm{e}^\alpha}{\mathrm{e}^\alpha + a - 1}, & \text{if } x = y, \\ \frac{1}{\mathrm{e}^\alpha + a - 1}, & \text{if } x \neq y. \end{cases} \tag{25}$$

A direct calculation then shows that

$$\mathrm{LIP}(\mathrm{GRR}^\alpha) = \max_{x,s} \left| \ln \frac{1 + (\mathrm{e}^\alpha - 1)\,\mathrm{p}_{x|s}}{1 + (\mathrm{e}^\alpha - 1)\,\mathrm{p}_x} \right|. \tag{26}$$

If we want GRR to satisfy ε-LIP, we then need to solve $\mathrm{LIP}(\mathrm{GRR}^\alpha) = \varepsilon$ 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

$$\mathbb{P}(E_{x'} = 1) = \begin{cases} \frac{1}{2}, & \text{if } x' = x, \\ \frac{1}{\mathrm{e}^\alpha + 1}, & \text{if } x' \neq x. \end{cases} \tag{27}$$

In other words, If we identify a $y \in 2^{\mathcal{X}}$ with a subset of $\mathcal{X}$ (so $\#y$ denotes its cardinality), we get

$$\mathrm{OUE}^\alpha_{y|x} = \begin{cases} \frac{\mathrm{e}^{(a - \#y)\alpha}}{2(\mathrm{e}^\alpha + 1)^{a-1}}, & \text{if } x \in y, \\ \frac{\mathrm{e}^{(a - \#y - 1)\alpha}}{2(\mathrm{e}^\alpha + 1)^{a-1}}, & \text{if } x \notin y. \end{cases} \tag{28}$$

It follows that

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

B. Conditional Reporting

In general, a generic LDP protocol will not be ideal for our situation, since these are designed to obscure all information about X, rather than just the part that holds information about S. To address this shortcoming, we introduce the *Conditional Reporting* (CR) 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, we return the true value of X. If not, we output a random one, whose probability distribution is given by $\mathrm{p}_{X|\tilde{s}}$.

Algorithm 1: Conditional Reporting (CR^α)

Input : Privacy parameter α; 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

$$\mathbb{P}(\tilde{S} = s') = \begin{cases} \frac{\mathrm{e}^\alpha}{\mathrm{e}^\alpha + \#\mathcal{S} - 1}, & \text{if } s' = s, \\ \frac{1}{\mathrm{e}^{\alpha_0} + \#\mathcal{S} - 1}, & \text{otherwise} \end{cases}$$

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, we can get better privacy, as outlined by the proposition below.

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

$$L(\alpha) = \max_{x,s} \left| \ln \frac{(e^\alpha - 1)\,\mathrm{p}_{x|s} + \sum_{s'} \mathrm{p}_{x|s'}}{(e^\alpha - 1)\,\mathrm{p}_x + \sum_{s'} \mathrm{p}_{x|s'}} \right|. \tag{30}$$

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* $L(\alpha) \leq \alpha$.

Proof. For all $y \in \mathcal{X}$ and $s \in \mathcal{S}$ we have, following equation (45) in Appendix A, that

$$\mathbb{P}(\mathrm{CR}^\alpha(X, S) = y | S = s) = \tfrac{1}{\mathrm{e}^\alpha + c - 1} \Big(\mathrm{e}^\alpha\, \mathrm{p}_{y|s} + \sum_{s' \neq s} \mathrm{p}_{y|s'} \Big). \tag{31}$$

It follows that

$$\begin{aligned} &\frac{\mathbb{P}(\mathrm{CR}^\alpha(X, S) = y | S = s)}{\mathbb{P}(\mathrm{CR}^\alpha(X, S) = y | S = s')} \\ &= \frac{\mathrm{e}^\alpha\, \mathrm{p}_{y|s} + p_{y|s'} + \sum_{s'' \neq s, s'} \mathrm{p}_{y|s''}}{\mathrm{p}_{y|s} + \mathrm{e}^\alpha p_{y|s'} + \sum_{s'' \neq s, s'} \mathrm{p}_{y|s''}} \qquad (32) \\ &\leq \max \left\{ 1, \frac{\mathrm{e}^\alpha\, \mathrm{p}_{y|s} + p_{y|s'}}{\mathrm{p}_{y|s} + \mathrm{e}^\alpha p_{y|s'}} \right\} \qquad (33) \\ &\leq \mathrm{e}^\alpha. \end{aligned}$$

□

VII. Experiments

We test the feasibility of the different methods 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

We compare the computing time for finding optimal ε-LDP and ε-LIP protocols 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. We take $\varepsilon \in \{0.5, 1, 1.5, 2\}$; the results are in Figure 3. 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.

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

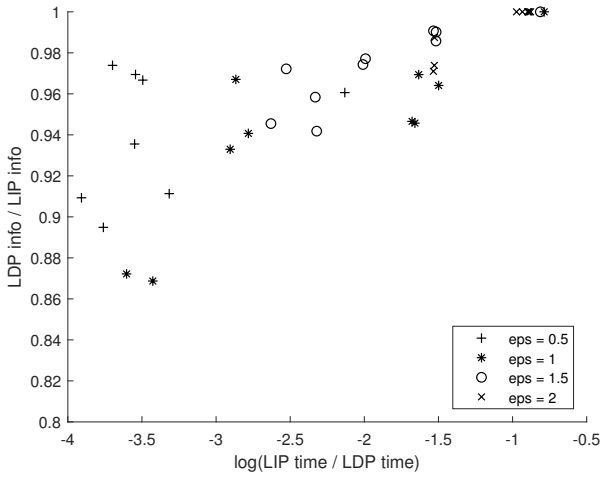


Figure 3. *Comparision of computation time and* $\mathrm{I}(X;Y)$ *for* ε*-LDP protocols found via Theorem 2 and* ε*-LIP protocols found via Theorem 3, for random* $\mathrm{p}_{S,X}$ *with* $c = 2$*,* $a = 5$*, and* $\varepsilon \in \{0.5, 1, 1.5, 2\}$.

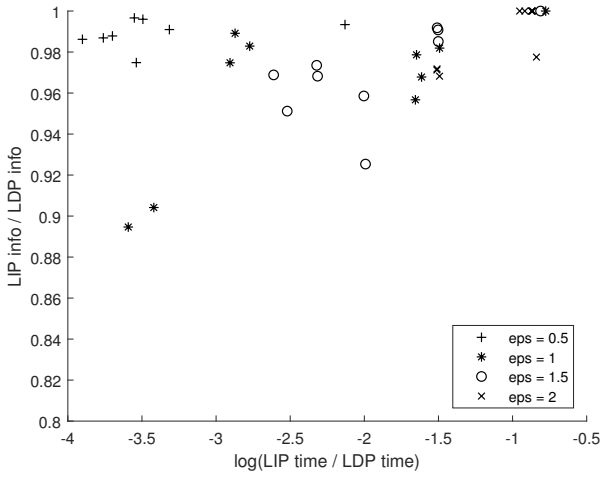


Figure 4. Comparision of computation time and $\mathrm{I}(X;Y)$ *for* ε*-LDP protocols found via Theorem 2 and* $\frac{\varepsilon}{2}$*-LIP protocols found via Theorem 3, for random* $\mathrm{p}_{S,X}$ *with* $c = 2$*,* $a = 5$*, and* $\varepsilon \in \{0.5, 1, 1.5, 2\}$.

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 4, we compare optimal $\frac{\varepsilon}{2}$-LDP protocols 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

We also perform similar comparisons for multiple attributes, for $c = 2$, $a_1 = a_2 = 3$ and $a_3 = 4$, comparing the methods of Theorems 3 and 4. The results are presented in Figure 5. As

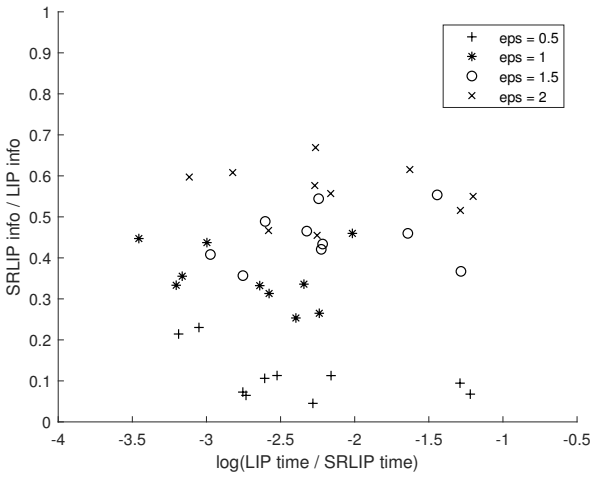


Figure 5. Comparison of computation time and $\mathrm{I}(X;Y)$ *for* ε*-(SR)LIP-protocols found via Theorems 3 and 4, for random* $\mathrm{p}_{S,X}$ *with* $c = 2$*,* $a_1 = a_2 = 3$*,* $a_3 = 4$*, and* $\varepsilon \in \{0.5, 1, 1.5, 2\}$.

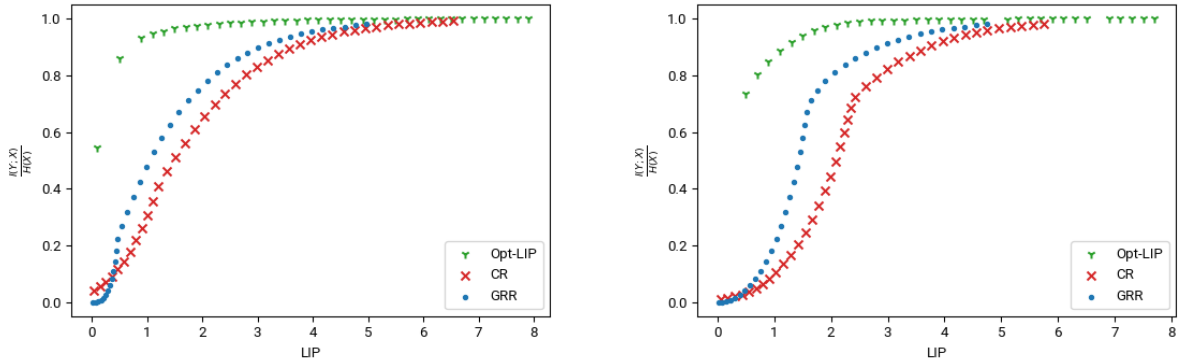
one can see, Theorem 4 is significantly slower, with Theorem 3 being on average 476 times as fast. There is a sizable difference in utility, caused on one hand by the fact that ε-SRLIP is a stricter privacy requirement than ε-LIP, and on the other hand by the fact that Theorem 4 does not give us the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

C. Adult dataset

We also test the utility of Conditional Reporting (CR), both on real world data and synthetic data. We consider the well-known Adult dataset [7], which contains demographic data from the 1994 US census. For our tests, we take $S \in \{\text{marital status, occupation}\}$ (with $c = 7$ and $c = 15$, respectively) and $X \in \{\text{education, relationship, sex}\}$ (with $a = 16, 6, 2$). Based on our findings in the previous sections, we take LIP as a privacy measure, and $\mathrm{I}(X;Y)$ as a utility measure. We compare CR on the one hand with the optimal method (Opt-LIP) found in Section IV, and on the other hand with the established LDP protocols GRR and OUE. The results are shown in Figure 6. 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 timeframe. Nevertheless, we can conclude 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 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.

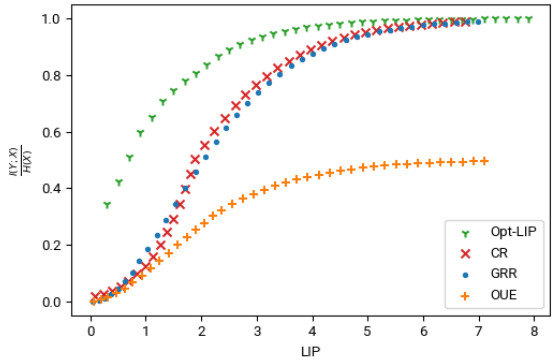
D. Synthetic data: GRR vs CR

To investigate the difference between GRR and CR, we apply both methods to synthetic data. We disregard OUE as it performs worse than the other two protocols, especially in

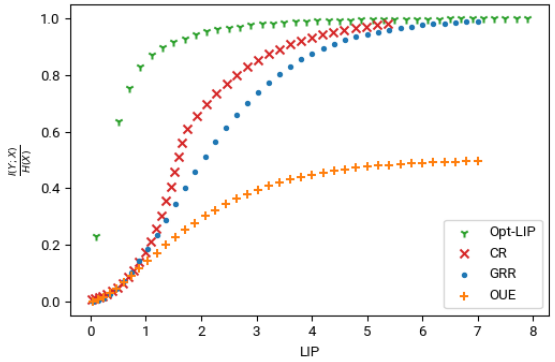


(a) $S =$ marital status, $X =$ education

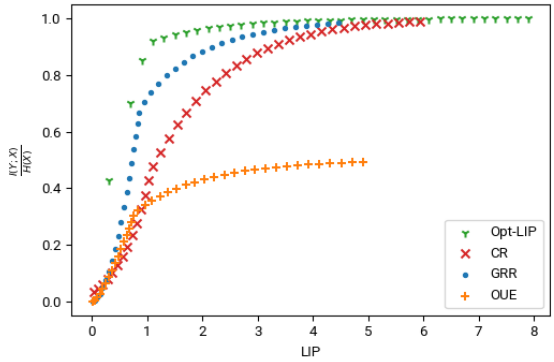
(b) $S =$ occupation, $X =$ education



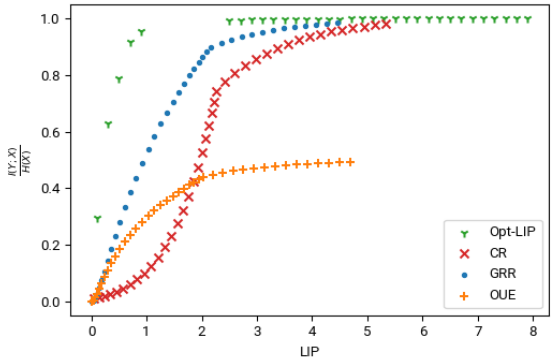
(c) $S =$ marital status, $X =$ relationship



(d) $S =$ occupation, $X =$ relationship



(e) $S =$ marital status, $X =$ race



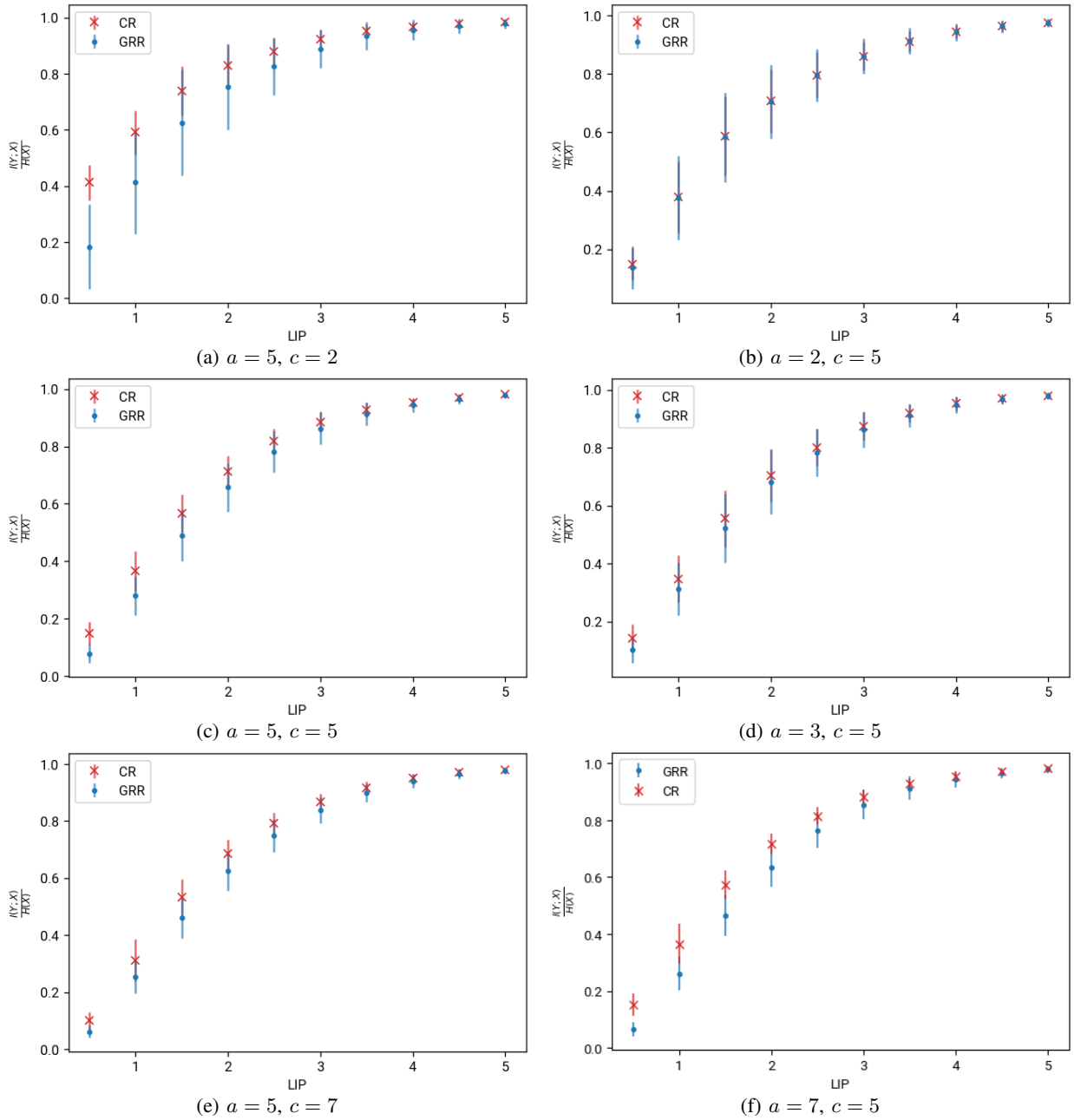
(f) $S =$ occupation, $X =$ race

Figure 6. *Experiments on the Adult dataset.*

the low privacy regime. For a fixed choice of a and c, we draw a number of probability distributions from the Jeffreys prior on $\mathcal{S} \times \mathcal{X}$, i.e. the symmetric Dirichlet distribution with parameter $\frac{1}{2}$. We fix a set of LIP values ε, and for each of these and each probability distribution, we solve equations (26) and (30), setting the left hand side equal to ε and solving for α_{GRR} and α_{CR}. We then calculate the mutual information $\mathrm{I}(X;Y)$, which we normalise by dividing by $\mathrm{H}(X)$. The resulting averages and standard deviations are displayed in Figure 7. On the whole, we see 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, we consider the parameters α_{CR} and α_{GRR} that govern the privacy protocols CR and GRR. 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



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

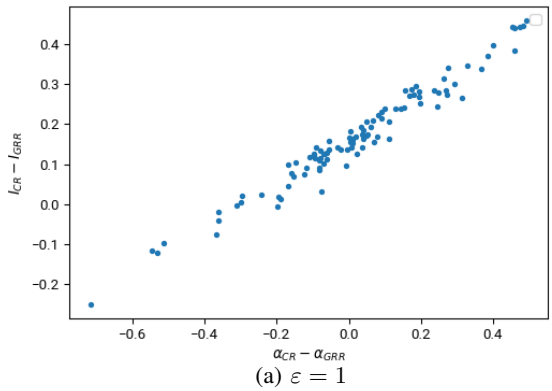
Figure 7. *Experiments on synthetic data. For each value of a and c, the average utility is taken over 100 randomly generated probability distributions. Bar size denotes standard deviation.*

probability distributions may be explained by a difference in α. We test this assertion for 100 randomly generated distributions in Figure 8. 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 9, we plot the relation between α and the LIP value ε for the experiments in 6(b) and 6(d). The fact that $\alpha_{\text{GRR}} > \alpha_{\text{CR}}$ in 9(a) corresponds to the fact that GRR outperforms CR in 6(b), and the opposite relation holds between 9(b) and 6(d).

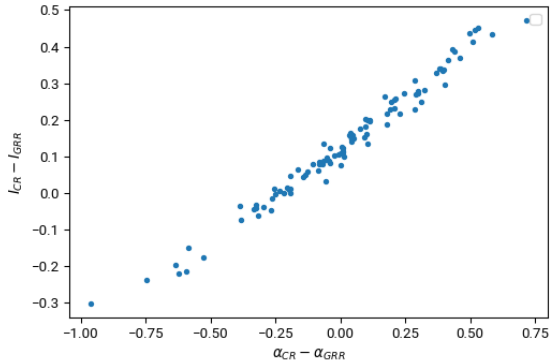
Unfortunately, we were not able to relate the difference in parameter α to other properties of the distribution. Without presenting details we mention that the properties $\mathrm{I}(X;S), \max_{x,s} \mathrm{p}_{x,s}, \max_x \mathrm{p}_x$ and $\max_s \mathrm{p}_s$ do not appear to have an impact on the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

VIII. Conclusions and Future Work

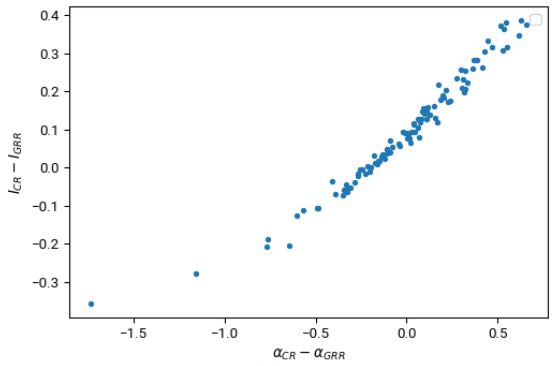
Local data sanitisation protocols have the advantage of being scalable for large numbers of users. Furthermore, the advantage of using differential privacy-like privacy metrics is that they provide worst-case guarantees, ensuring that the



(a) $\varepsilon = 1$



(b) $\varepsilon = 1.5$



(c) $\varepsilon = 2$

Figure 8. *Difference in* α *versus difference in utility for* 100 *randomly generated probability distributions, for* $a = c = 5$.

privacy of every user is sufficiently protected. For both ε-LDP and ε-LIP we have derived methods to find optimal sanitisation protocols. Within this setting, we have observed that ε-LIP has two main advantages over ε-LDP. First, it fits better within the privacy funnel 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, we have 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 we pay for this both in computation time and in utility.

With regard to the specific protocols, we have found that the newly introduced protocol, CR, generally outperforms OUE, especially for high values of ε-LIP. It 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, we have seen that CR performs better on average if a is large compared to c.

For further research, a number of important avenues remain 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 [3], 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.

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

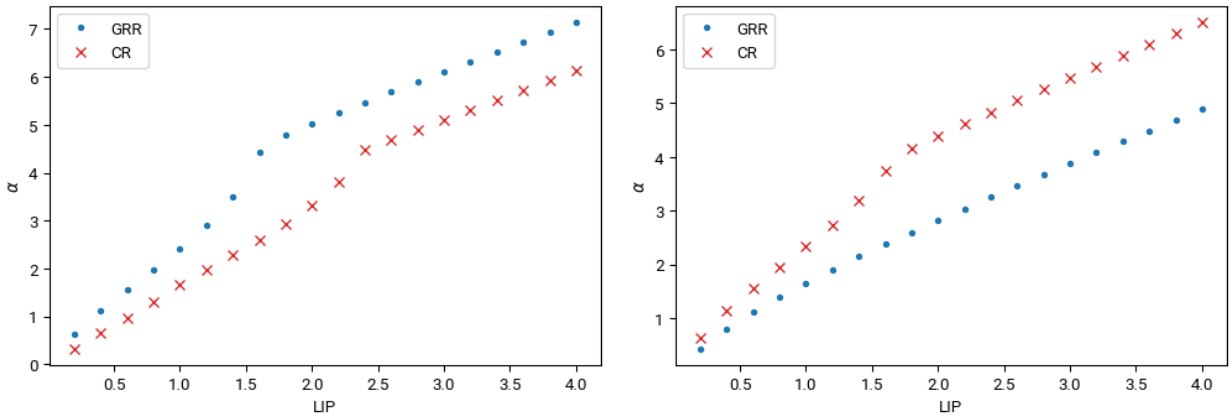
Acknowledgements

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

References

[1] 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.

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



(a) $S =$ occupation, $X =$ education (b) $S =$ occupation, $X =$ relationship

Figure 9. *Value of GRR and CR parameter α for different values of ε for the Adult dataset.*

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

[4] Flavio du Pin Calmon et al. "Principal inertia components and applications". In: *IEEE Transactions on Information Theory* 63.8 (2017), pp. 5011–5038.

[5] 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.

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

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

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

[9] Cynthia Dwork et al. "Calibrating noise to sensitivity in private data analysis". In: *Theory of cryptography conference*. Springer. 2006, pp. 265–284.

[10] Cynthia Dwork et al. "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.

[11] Úlfar Erlingsson et al. "Encode, shuffle, analyze privacy revisited: formalizations and empirical evaluation". In: *arXiv:2001.03618* (2020).

[12] 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.

[13] 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.

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

[15] Shiva Prasad Kasiviswanathan et al. "What can we learn privately?" In: *SIAM Journal on Computing* 40.3 (2011), pp. 793–826.

[16] 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.

[17] 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.

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

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

[20] Ali Makhdoumi et al. "From the information bottleneck to the privacy funnel". In: *2014 IEEE Information Theory Workshop (ITW 2014)*. IEEE. 2014, pp. 501–505.

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

[22] Fabian Prasser et al. "Arx-a comprehensive tool for anonymizing biomedical data". In: *AMIA Annual Symposium Proceedings*. Vol. 2014. American Medical Informatics Association. 2014, p. 984.

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

[24] Salman Salamatian et al. "Privacy-Utility Tradeoff and Privacy Funnel". In: (2020). unpublished preprint. URL: http://www.mit.edu/~salmansa/files/privacy_TIFS.pdf.

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

[26] Tianhao Wang et al. "Locally differentially private protocols for frequency estimation". In: *26th {USENIX} Security Symposium ({USENIX} Security 17)*. 2017, pp. 729–745.

[27] 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.

[28] Mengmeng Yang et al. "Local Differential Privacy and Its Applications: A Comprehensive Survey". In: *arXiv:2008.03686* (2020). Preprint.

APPENDIX A
PROOFS

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

$$\begin{aligned}
\mathrm{p}_{y|s,x^J} &= \sum_{x^{\setminus J}} \mathrm{p}_{y|x}\, \mathrm{p}_{x^{\setminus J}|s,x^J} && (34)\\
&= \mathrm{p}_{y^J|x^J} \sum_{x^{\setminus j}} \left(\prod_{j\notin J} \mathrm{p}_{y^j|x^j} \right) \mathrm{p}_{x^{\setminus J}|s,x^J} && (35)\\
&= \mathrm{p}_{y^J|x^J} \sum_{x^{\setminus j}} \prod_{j\notin J} \mathrm{p}_{y^j|x^j}\, \mathrm{p}_{x^j|s,x^{J[j]}} && (36)\\
&= \mathrm{p}_{y^J|x^J} \prod_{j\notin J} \sum_{x^j} \mathrm{p}_{y^j|x^j}\, \mathrm{p}_{x^j|s,x^{J[j]}} && (37)\\
&= \mathrm{p}_{y^J|x^J} \prod_{j\notin J} \mathrm{p}_{y^j|s,x^{J[j]}} && (38)\\
&\le \mathrm{p}_{y^J|x^J} \prod_{j\notin J} \mathrm{e}^{\varepsilon^j}\, \mathrm{p}_{y^j|x^{J[j]}} && (39)\\
&\le \mathrm{e}^{\varepsilon}\, \mathrm{p}_{y^J|x^J} \prod_{j\notin J} \mathrm{p}_{y^j|x^{J[j]}} && (40)\\
&= \mathrm{e}^{\varepsilon}\, \mathrm{p}_{y|x^J}\,. && (41)
\end{aligned}$$

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

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

$$\begin{aligned}
Q_{y|x,s} &= \sum_{s'} \mathbb{P}(\mathrm{CR}^\alpha(x,s) = y|\tilde{s} = s')\mathbb{P}(\tilde{s} = s'|S = s) && (42)\\
&= \frac{\mathrm{e}^\alpha}{\mathrm{e}^\alpha + c - 1} + \frac{1}{\mathrm{e}^\alpha + c - 1} \sum_{s'\neq s} \mathrm{p}_{y|s'}, && (43)
\end{aligned}$$

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

$$\begin{aligned}
&\mathbb{P}(\mathrm{CR}^\alpha(X,S) = y|S = s)\\
&= \sum_x Q_{y|x,s}\, \mathrm{p}_{x|s} && (44)\\
&= \frac{\mathrm{e}^\alpha}{\mathrm{e}^\alpha + c - 1}\, \mathrm{p}_{y|s} + \frac{1}{\mathrm{e}^\alpha + c - 1} \sum_{s'\neq s} \mathrm{p}_{y|s'} && (45)\\
&= \frac{\mathrm{e}^\alpha - 1}{\mathrm{e}^\alpha + c - 1}\, \mathrm{p}_{y|s} + \frac{1}{\mathrm{e}^\alpha + c - 1} \sum_{s'} \mathrm{p}_{y|s'}, && (46)\\
&\mathbb{P}(\mathrm{CR}^\alpha(X,S) = y)\\
&= \sum_s \mathbb{P}(\mathrm{CR}^\alpha(X,S) = y|S = s)\, \mathrm{p}_s && (47)\\
&= \frac{\mathrm{e}^\alpha}{\mathrm{e}^\alpha + c - 1}\, \mathrm{p}_y + \frac{1}{\mathrm{e}^\alpha + c - 1} \sum_s \sum_{s'\neq s} \mathrm{p}_{y|s'}\, \mathrm{p}_s && (48)\\
&= \frac{\mathrm{e}^\alpha}{\mathrm{e}^\alpha + c - 1}\, \mathrm{p}_y + \frac{1}{\mathrm{e}^\alpha + c - 1} \sum_{s'} \mathrm{p}_{y|s'} \sum_{s\neq s'} \mathrm{p}_s && (49)\\
&= \frac{\mathrm{e}^\alpha}{\mathrm{e}^\alpha + c - 1}\, \mathrm{p}_y + \frac{1}{\mathrm{e}^\alpha + c - 1} \sum_{s'} (\mathrm{p}_{y|s'} - \mathrm{p}_{y,s'}) && (50)\\
&= \frac{\mathrm{e}^\alpha - 1}{\mathrm{e}^\alpha + c - 1}\, \mathrm{p}_y + \frac{1}{\mathrm{e}^\alpha + c - 1} \sum_{s'} \mathrm{p}_{y|s'}\,. && (51)
\end{aligned}$$

We find that

$$L(\alpha) = \max_{y,s} \left| \ln \frac{\mathbb{P}(\mathrm{CR}^\alpha(X,S) = y|S = s)}{\mathbb{P}(\mathrm{CR}^\alpha(X,S) = y)} \right|, \qquad (52)$$

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