Local Information Privacy and Its Application to Privacy-Preserving Data Aggregation

Bo Jiang, Student Member, IEEE, Ming Li, Senior Member, IEEE, and Ravi Tandon, Senior Member, IEEE

Abstract—In this paper, we study local information privacy (LIP), and design LIP based mechanisms for statistical aggregation while protecting users’ privacy without relying on a trusted third party. The notion of context-awareness is incorporated in LIP, which can be viewed as explicit modeling of the adversary’s background knowledge. It enables the design of privacy-preserving mechanisms leveraging the prior distribution, which can potentially achieve a better utility-privacy tradeoff than context-free notions such as Local Differential Privacy (LDP). We present an optimization framework to minimize the mean square error in the data aggregation while protecting the privacy of each individual user’s input data or a correlated latent variable while satisfying LIP constraints. Then, we study two different types of applications: (weighted) summation and histogram estimation, and derive the optimal context-aware data perturbation parameters for each case, based on randomized response type of mechanism. We further compare the utility-privacy tradeoff between LIP and LDP and theoretically explain why the incorporation of prior knowledge enlarges feasible regions of the perturbation parameters, which thereby leads to higher utility. We also extend the LIP-based privacy mechanisms to the more general case when exact prior knowledge is not available. Finally, we validate our analysis by simulations using both synthetic and real-world data. Results show that our LIP-based privacy mechanism provides better utility-privacy tradeoffs than LDP, and the advantage of LIP is even more significant when the prior distribution is more skewed.

Index Terms—privacy-preserving data aggregation, local information privacy, information-theoretic privacy

1 INTRODUCTION

Privacy issues are crucial in this big data era, as users’ data are collected both intentionally or unintentionally by an increasing number of private or public organizations. Most of the collected data is used for ensuring high quality of service, but may also put one’s sensitive information at potential risk. For instance, when someone is rating a movie, his/her preferences may be leaked; when someone is searching for a parking spot nearby using a smartphone, his/her real location is uploaded and prone to leakage. Besides the cases where collected data itself is sensitive and causes privacy leakage, non-sensitive data release may also enable malicious inference on one’s private attributes: whenever there is a correlation between the collected data and people’s private latent attribute, directly releasing it causes privacy leakage. For instance, heartbeat data collected by smartwatch may potentially reveal one’s heart disease [1]. One can easily infer a target user’s home or work location by tracking his daily location data [2]; Smart meters can reveal the activities of people inside a home by tracking their electricity, gas, or water usage frequently over time [3]. It is therefore desirable to design privacy-preserving mechanisms providing privacy guarantees without affecting data utility.

Traditional privacy notions such as $k$-anonymity [4] do not provide rigorous privacy guarantees and are prone to various attacks. Nowadays, Differential Privacy (DP) [5], [6] has become the de facto standard for ensuring data privacy in the database community [7]. The definition of DP assures that each user’s data has minimal influence on the output of statistical queries on a database. In the classical DP setting, a trusted server is assumed to hold all users’ data and provide noisy answers to queries. However, organizations or companies collecting users’ data may not be trustworthy, and the data storage system may not be secure [8]. As a result, recently, local privacy protection mechanisms have gained attention as the local setting allows data aggregation while protecting each user’s data without relying on a trusted third party [9].

1.1 Local Privacy Notions

In local privacy-preserving data release, each user perturbs his or her data locally before uploading it; organizations that want to take advantage of users’ data then aggregate over the collected data. The earliest such mechanism is randomized response [10], which randomly perturbs each user’s data. However, the original randomized response does not have formal privacy guarantees. Later, Local Differential Privacy (LDP) was proposed as a local variant of DP that bounds the privacy leakage in the local setting [11]. Many schemes were proposed under the notion of LDP. For example, [12]–[14], and Google’s RAPPOR [15]. LDP based data aggregation mechanisms have already been deployed in the real-world. For example, in June 2016, Apple announced that it would deploy LDP-based mechanisms to collect user’s typing data [16]. However, Tang et al. show that although each user’s perturbation mechanism satisfies LDP, the privacy budget is too large ($\epsilon = 43$) to provide any useful privacy protection. Wang et al. proposed a variety of

1. The parameters, $\epsilon \geq 0$, measures the privacy level. A smaller $\epsilon$ corresponds to a higher privacy level.
LDP protocols for frequency estimation [17] and compared their performance with Google’s RAPPOR. However, for a given reasonable privacy budget, these protocols provide limited utility. Intuitively, compared with the central DP model, it is more challenging to achieve a good utility-privacy tradeoff in the local setting. The main reasons are twofold: (1) LDP requires introducing noise at a significantly higher level than what is required in the central setting. That is, a lower bound of noise magnitude of $\Omega(\sqrt{N})$ is required for LDP, where $N$ is the number of users. In contrast, only $O(1)$ is required for central DP [18]. (2) LDP does not assume a neighborhood constraint on input data, thus when the data domain is large, LDP leads to a significantly reduced utility [19].

In general, both local and central DP provide strong context-free theoretical guarantees against worst-case adversaries [20]. Context-free means the adversary can possess arbitrary background knowledge of a user’s data (except her specific input instance). In other words, the definition of (L)DP is too strong and regardless of scenarios where particular context or prior knowledge of the data is available. Such scenarios exist for many applications, for instance, in Internet of Things (IoT), distribution of context related to sensor data plays critical role in distributed data transmission and computation [21]; another example is location based services, people have a higher likelihood to be at some locations than others; such as in Pairs, people are more likely closer to Eiffel tower than a coffee shop nearby [22]. In mobile-health data collection, background knowledge such as the likelihood of people having certain diseases is available through previously published medical studies [23]. When background information is available, (L)DP fails to capture the explicit privacy leakage of user or the information gain at the adversary. On the other hand, for a given utility, (L)DP may not always be feasible depending on the privacy budget [24]. Although approximate $(\epsilon, \delta)$-(L)DP is introduced [25], [26] to realize an achievable mechanism, the non-negative addend $\delta$ could be large enough (close to 1) to provide limited privacy guarantee.

1.2 Relaxing Local Differential Privacy

There’s a trend among the privacy research community that leverages the background knowledge to relax the definition of DP and the utility can be increased by explicitly modeling the adversary’s knowledge. Privacy notions that consider such prior knowledge are denoted as “context-aware” privacy notions. For context-aware privacy notions, besides the privacy budget $\epsilon$, the amount of required noise also depends on the prior distribution of the data: context-dependent privacy mechanisms add noise selectively according to the data prior when most needed so that utility can be enhanced. For example, less noise is required to perturb for data with higher certainty [20], [27]. In general the existing context-aware privacy definitions fall into two categories based on either average-case or worst-case guarantees. All information-theoretic privacy notions belong to the former class [28]–[30]. The latter includes Pufferfish [31], Bayes DP [32], Membership privacy [33], etc. Average notions are generally weaker than the latter since they cannot bound the leakage for all the input and output pairs, which may not be easily adopted by the users who are privacy-sensitive. On the other hand, existing context-aware worst-case privacy notions like Pufferfish and Bayes DP still follow the same structure of (L)DP – maximum ratio between two likelihoods of a certain output given different input data. Since the relationship with prior distribution is not directly captured in the privacy notion, this makes context-aware privacy mechanism design very difficult (either high complexity or not easily composable).

1.3 Local Information Privacy

In this paper, we make use of the maximum ratio of posterior to prior to capture information leakage in the local setting, denote as local information privacy (LIP). Originally, information privacy (IP) was proposed in a central setting by Calmon et. al. [34], which requires a trusted curator. Another reason that prohibits Centralized IP from being adopted in practice is that the distribution of all users’ data is too complex to express or capture, especially for a large-size dataset. In contrast, LIP requires only the prior distribution of one particular user’s data, which can be obtained through many approaches in practice.

An illustrative example of why context-aware privacy notions result in increased utility is shown in Fig.1, which shows the perturbation mechanisms of context-free (LDP) and context-aware (LIP) notions and the comparison of the mean square errors when collecting private binary data with specific prior. We illustrate the optimal perturbation probabilities for the same privacy budget (epsilon=0.6) under both LDP and LIP privacy notions. Observe that the perturbation channel of LDP is symmetric, while LIP designs perturbation parameters according to the prior knowledge. When the data value is quite certain, it has smaller probability to flip the value in order to increase the utility, while when the data takes a value that has small probability of happening, the mechanism also protect its privacy by a large perturbation probability (a large amount of additive noise). In this example, the probability of flipping the data value through the LDP mechanism is 0.35 in contrast to $0.2 \times 0.55 + 0.8 \times 0.1 = 0.19$ of the LIP mechanism. As a result, LIP leads to an enhanced utility than LDP.

1.4 Main Contributions

The main contributions of this paper are listed as follow:

(1) We study a privacy notion called Local Information Privacy (LIP) for local data release (without a trusted third party), which relaxes the notion of LDP by incorporating prior knowledge. We formally derive the relationships between existing definitions and LIP.

(2) We apply the LIP notion to privacy-preserving data aggregation: we present a general framework to estimate a function of the collected data and minimize the mean squared error of the estimation while protecting each individual’s privacy by satisfying LIP constraints. Our mechanism can be viewed as a general form of randomized response, also, we derive the optimal perturbation parameters and the utility-privacy tradeoff. We further consider a latent variable model, where the collected data is correlated with some private latent properties of each user.
(3) We show how the proposed framework can be applied to different applications, including weighted summation as well as histogram. Both the optimal estimator at the curator and the perturbation mechanism at the user are derived. We derive the closed-form solutions for the model with the same utility function but with LDP privacy constraints, and theoretically compare the utility-privacy tradeoff with LIP models.

(4) We validate our analysis using simulations on both synthetic and real-world datasets (Karosak, a website-click stream data set, and Adult, a survey of census income). Both theoretical and simulation results show that mechanisms under $\epsilon$-LIP always achieve a better utility-privacy tradeoff than those under $\epsilon$-LDP.

Next, we enumerate the main differences from the version in IEEE CNS 2018 [35].

1. The system model and framework generalization: In [35], the aggregated data is binary and it is private, in this journal paper, besides generalizing to the setting where the data domain contains multiple values, we also consider a model where a private latent variable is correlated with the aggregated data, such that the utility is measured by the error in estimating the aggregated function, while protecting the privacy of the latent variable. (2) We generalize the privacy definition and explore its relationships with existing definitions. (3) Theoretical analysis enhancement: In the conference version, we derive the closed-form optimization parameters under the binary model. In this journal extension, the utility-privacy tradeoff under the MIMO model is derived. Also, we extend the model and analysis to consider inaccurate/bounded prior knowledge. Additionally, for the latent variable model, we show that given a privacy budget, there is a prior distribution dependent threshold under which the optimal mechanism directly releases the raw data. (4) Realizations of more general applications: the binary model in the conference version is only suitable for the application of survey. In this journal extension, we use the MIMO framework to handle more general applications besides survey, including (weighted) summation and histogram. (5) Extensive experimental validation: Besides comparing the performance of the proposed mechanism with binary data, we make the following extensions in experiments: a). We generate synthetic data and use real-world data with a large domain and study the effect of data domain size on the utility-privacy tradeoff. b). Machine learning methods are deployed to learn the correlation between the aggregated data and latent variables using real-world datasets. c). We compare with both context-aware LDP and context-free LDP to illustrate the need for introducing contextual information.

The remainder of the paper is organized as follows. In Section 2, we describe the proposed LIP notion and its relationship with other existing privacy notions. In Section 3, we introduce the system model and problem formulation. In Section 4, we derive the utility-privacy tradeoff under different models including MIMO model, model with uncertain prior and comparison with LDP model, and then we discuss applications of these models. In Section 5, the model with private latent variable is presented. In Section 6 we present the simulation results and compare utility-privacy tradeoffs for a variety of data sets and applications. In Section 7, we offer concluding remarks.

2 PRIVACY DEFINITIONS AND RELATIONSHIPS

In this Section, we first recap several existing privacy notions in the local settings, we then extend Information Privacy to the local setting (LIP); finally, we study its relationships with other notions.

2.1 Privacy Definitions

Denote the raw private data as $X$, which takes value from domain $\mathbb{D}$, and has a prior $\theta_X$. A privacy-preserving mechanism $M$ takes $X$ as input, and releases a perturbed version of $Y$ as an output. Denote $Y = \text{Range}(M)$ as the domain of $Y$.

The context-free LDP definition states that any two inputs from the data domain $\mathbb{D}$ result in the same output with similar probabilities.

Definition 1. ($\epsilon$-Local Differential Privacy (LDP)) [36] A mechanism $M$ which takes input $X$ and outputs $Y$ satisfies $\epsilon$-LDP for some $\epsilon \in \mathbb{R}^+$, if $\forall x, x' \in \mathbb{D}$ and $\forall y \in Y$:

$$\frac{Pr(Y = y | X = x)}{Pr(Y = y | X = x')} \leq e^\epsilon.$$  

(1)

LDP provides strong privacy protection, since it guarantees indistinguishability of user’s inputs regardless of the data prior distribution, and has led to multiple real-world applications. However, such strong (worst-case) privacy definition also leads to a significant utility loss.

When some context of $X$ is available (e.g., prior distribution), one can incorporate context (prior) information into the privacy definition to explicitly model the adversaries’ knowledge. One such definition is pufferfish privacy, where the adversary’s knowledge is defined by a subset of all possible prior distributions. Pufferfish privacy is originally proposed in the central setting [31], here we adapt it into
the local setting where \( X \) and \( Y \) are each user's input and output, respectively.

**Definition 2** (Local Pufferfish Privacy). Given a set of potential secrets \( G \), a set of discriminative pairs \( G_{\text{pairs}} \), a set of data evolution scenarios \( \mathcal{P} \), a privacy protection mechanism \( \mathcal{M} \) satisfies \( \epsilon \)-Pufferfish \((G, G_{\text{pairs}}, \mathcal{P}) \) privacy, for some \( \epsilon \in \mathbb{R}^+ \) if

- for all possible inputs \( x \in \mathbb{D}, y \in \mathcal{Y} \),
- for all pairs \((g_i, g_j) \in G_{\text{pairs}}\) of potential secrets,
- for all distributions \( \theta \in \mathcal{P} \) for which \( \Pr(g_i(\theta) \neq 0 \) and \( \Pr(g_j(\theta) \neq 0 \)

the following holds:

\[
e^{-\epsilon} \leq \frac{\Pr(\mathcal{M}(x) = y|\theta, g_i)}{\Pr(\mathcal{M}(x) = y|\theta, g_j)} \leq e^\epsilon.
\] (2)

In the definition of Pufferfish above, the raw data \( X \) is correlated with a private latent variable \( G \), and the distribution \( \theta \) captures their joint distribution. When the set of priors \( \mathcal{P} \) spans all possible joint distributions, this also includes the case when \( X = G \), and for such a special case, pufferfish becomes equivalent to LDP. More generally, however, pufferfish is a relaxation of LDP. Despite being able to introduce priors in the privacy definition, one of the drawbacks of Pufferfish privacy is the difficulty of mechanism design. Recently, in [37], Wang et al. designed a Wasserstein Mechanism, which achieves Pufferfish privacy, but it is computationally inefficient, and the mechanism they proposed is approximated.

Another context-aware privacy definition is mutual information privacy, which uses the mutual information between \( Y, X \) to measure the average information leakage of \( X \) contained in \( Y \):

**Definition 3.** (\( \epsilon \)-Mutual Information Privacy (MIP)) [29] A mechanism \( \mathcal{M} \) which takes input \( X \) and outputs \( Y \), satisfies \( \epsilon \)-MIP for some \( \epsilon \in \mathbb{R}^+ \), if the mutual information between \( X \) and \( Y \) satisfies

\[
\sum_{x \in \mathbb{D}, y \in \mathcal{Y}} \Pr(X = x, Y = y) \log \frac{\Pr(X = x, Y = y)}{\Pr(Y = y) \Pr(X = x)} \leq \epsilon, \text{ where } I(X; Y) =
\] (3)

Although MIP is context-aware, it provides relative weak privacy protection since it only bounds the average information leakage over all possible \( x \) and \( y \) in the domain. There may exist some \( (x, y) \) pairs making the ratio between the joint and product of marginal distributions very large (while the joint probability is very small).

Maximal Information Leakage, as a stronger privacy notion than MIP, is defined as:

**Definition 4.** (\( \epsilon \)-Maximal Information Leakage (MIL)) [33] For a mechanism \( \mathcal{M} \) which takes input \( X \) and outputs \( Y \), the maximal information leakage is defined as:

\[
\mathcal{L}(X \rightarrow Y) = \log \sum_{y \in \mathcal{Y}} \max_{x \in \mathbb{D}} \Pr(Y = y|X = x),
\] (4)

and \( \mathcal{M} \) satisfies Maximal Information Leakage privacy if for some \( \epsilon \in \mathbb{R}^+, \forall x \in \mathbb{D}, y \in \mathcal{Y} \): \( \mathcal{L}(X \rightarrow Y) \leq \epsilon \).

MIL captures the average likelihood over all possible \( y \in \mathcal{Y} \) given the corresponding value of \( x \) that provides the maximal likelihood probability. However, MIL does not provide pairwise protection over all possible values of \( x \) and \( y \) and hence is relatively weak.

Another context-aware privacy definition that provides pairwise protection over each possible values of \( x \) and \( y \) is differential identifiability.

**Definition 5.** (\( \epsilon \)-Differential Identifiability (DI)) [39] A mechanism \( \mathcal{M} \) which takes input \( X \) and outputs \( Y \) satisfies \( \epsilon \)-DI for some \( \epsilon \in \mathbb{R}^+ \), if for all \( x, x' \in \mathbb{D} \) and \( \forall y \in \mathcal{Y} \):

\[
\frac{\Pr(X = x|Y = y)}{\Pr(X = x'|Y = y)} \leq e^\epsilon.
\] (5)

The operational meaning of DI is, given the output \( y \), the adversary cannot tell whether the original data set is \( x \) or \( x' \). DI can be directly adapted for the local setting, and is context-aware because DI also depends on the prior distribution:

\[
\frac{\Pr(Y = y|X = x)}{\Pr(Y = y|X = x')} \leq e^\epsilon.
\]

Observe that, the likelihood of \( y \) given \( x \) (the perturbation parameters) depends on the prior of \( x \). One major drawback of DI is the difficulty of designing practical mechanisms, as DI is based on the ratio of posteriors, after transferring to the ratios of two likelihoods, the likelihoods (perturbation parameters) depend on the priors. For example, if \( \Pr(X = x) \) is small, DI requires \( \Pr(Y = y|X = x')/\Pr(Y = y|X = x) \) to be large for all \( y \in \mathcal{Y} \). However, we know that \( \sum_{y \in \mathcal{Y}} \Pr(Y = y|X = x) = 1 \).

To provide a pairwise constraint on the information leakage of \( X \) through \( Y \), we consider a bound on the ratio between the prior and posterior, which leads to the notion of local information privacy:

**Definition 6.** (\( \epsilon \)-Local Information Privacy (LIP)) Given a set of potential prior distributions \( \mathcal{P}_X \), a mechanism \( \mathcal{M} \) which takes input \( X \) and outputs \( Y \) satisfies \( \epsilon \)-LIP for some \( \epsilon \in \mathbb{R}^+ \), if for all \( x \in \mathbb{D}, \forall \theta_X \in \mathcal{P}_X \) and \( \forall y \in \mathcal{Y} \):

\[
e^{-\epsilon} \leq \frac{\Pr(X = x|\theta_X)}{\Pr(X = x|Y = y, \theta_X)} \leq e^{\epsilon}.
\] (6)

There are three cases regarding the range of \( \mathcal{P}_X \):

- When \( \mathcal{P}_X \) includes one given prior distribution, LIP becomes LIP for fixed prior \( \theta_X \);
- When \( \mathcal{P}_X \) includes all possible priors, LIP becomes Worst-Case-LIP (WC-LIP);
- When \( \mathcal{P}_X \) includes a subset of all possible priors, LIP becomes Bounded-Prior-LIP (BP-LIP).

Intuitively, LIP guarantees that having the knowledge of the prior distribution, the adversary cannot infer too much additional information about each input \( x \) by observing each output \( y \). Note that, when \( \epsilon \) is small, this ratio is bounded close to 1.

The operational meaning of LIP can be interpreted as protecting a user's data privacy given the set of all the possible prior distribution(s) \( \mathcal{P}_X \). Compared with other context-aware definitions, LIP (including BP-LIP and WC-LIP) provides comprehensive modeling of the prior attainability, including the scenarios where the prior knowledge is uncertain, (WC-LIP can be viewed as context-free).
2.2 Relationships with Existing Definitions

2.2.1 LIP v.s. LDP

The following relationship holds between LIP and LDP: \( \epsilon \)-LIP implies \( 2\epsilon \)-LDP and \( \epsilon \)-LDP implies \( \epsilon \)-LIP (proof is shown in [35]). This implies that \( \epsilon \)-LIP is a more relaxed privacy notion than \( \epsilon \)-LDP; however, it is stronger than \( 2\epsilon \)-LDP. Intuitively, LIP relaxes LDP because LDP defenses against adversaries with arbitrary data prior v.s. a fixed prior of LIP.

When comparing the relationship between \( \epsilon \)-WC-LIP and \( \epsilon \)-LDP, we have \( \epsilon \)-WC-LIP is equivalent to \( \epsilon \)-LDP (proof is shown in [40]). Intuitively, these two definitions are equivalent because both of them assume worst-case (context-free) priors. Then the relationship between LDP and BP-LIP is straightforward: \( \epsilon \)-BP-LIP is sandwiched between \( \epsilon \)-LDP and \( \epsilon \)-LIP. As a result, LIP (BP-LIP, WC-LIP) can be viewed as context-aware versions of LDP with respect to different assumptions on the data priors. We further compare the utility privacy tradeoff between these two definitions in terms of optimal mechanism design in Sec. 4.3.

2.2.2 LIP v.s. Local Pufferfish

Since in the definition of Local Pufferfish, \( X \) is correlated with latent variable \( G \), to make a fair comparison with LIP, we consider the same situation, where \( Y \) is a noisy version of \( X \), which is correlated with a private latent variable \( G \). Denote \( \theta_{XG} \in \mathcal{P}_{XG} \) as the joint distribution of \( X \) and \( G \), where \( \mathcal{P}_{XG} \) is the set of possible joint distributions of \( X, G \). Then, the definition of \( \epsilon \)-LIP with latent variable for some \( \epsilon \in \mathbb{R}^+ \) can be written as:

\[
\forall \theta_{XG} \in \mathcal{P}_{XG}, \forall y \in G \text{ and } \forall y \in \mathcal{Y}: \\
\epsilon^{-1} \leq \frac{Pr(G = g|\theta_{XG})}{Pr(G = g|Y = y, \theta_{XG})} \leq \epsilon. \tag{7}
\]

Considering the same set of potential joint distributions (\( P \) in Local Pufferfish is identical to the \( \mathcal{P}_{XG} \)) in the two privacy notions, we next compare LIP (BP-LIP, WC-LIP) with Local Pufferfish privacy according to different scenarios of \( \mathcal{P}_{XG} \). The results in the next lemma follow from the proof of the relationship between LIP and LDP.

**Lemma 1.** The relationship between \( \epsilon \)-LIP and \( \epsilon \)-Local Pufferfish can be described as follow:

- \( \epsilon \)-WC-LIP is equivalent to \( \epsilon \)-Local Pufferfish when \( \mathcal{P}_{XG} \) includes all possible prior distributions of \( X \) and \( G \);
- \( \epsilon \)-Local Pufferfish implies \( \epsilon \)-BP-LIP, and \( \epsilon \)-BP-LIP implies \( 2\epsilon \)-Local Pufferfish when \( \mathcal{P}_{XG} \) includes a subset of all possible prior distributions of \( X \) and \( G \).

When \( \mathcal{P}_{XG} \) includes all possible prior distributions of \( X \) and \( G \), \( \epsilon \)-Local Pufferfish considers \( X = G \) (where the leakage is maximized), which is equivalent to \( \epsilon \)-LDP.

In summary, Local Pufferfish relaxes LDP by defining a bounded set of possible prior distributions. Since the “differential structure” in the definition of LDP does not allow for the incorporation of prior knowledge, Pufferfish further extends it by a correlated latent variable. This definition is more general in terms of operational meaning. However, it also comes with difficulties in mechanism design, as the values of \( Pr(Y = y|G = g) \) averages over all the likelihood probabilities of \( Pr(Y = y|X = x) \), which are the perturbation parameters. On the other hand, LIP relaxes LDP without the “differential structure” in the definition, thus the prior knowledge can be directly leveraged in the mechanism. We further study the model with latent variable in Sec. 5.

2.2.3 LIP v.s. Other Context-aware Notions

We next compare the relationship between LIP and other context-aware privacy notions under the assumption that \( \mathcal{P}_X \) contains only the exact prior of \( X \). Then, we have LIP is stronger than MIP, since \( Pr(X = z|Y = y) = Pr(X = z|Y = y) \leq \epsilon^\epsilon \). LIP is also stronger than MIL, as \( \max_{\epsilon \in \mathbb{D}} Pr(Y = y|X = x) \leq Pr(Y = y)e^\epsilon \). Intuitively, among LIP, MIP and MIL, only LIP provides pairwise protection over each possible realization of \( x \) and \( y \).

To compare the relationship between LIP and DI, we first define the maximal ratio of two prior probabilities of \( X \) as \( D_X^\epsilon = \max_{x, x' \in \mathbb{D}} \log \frac{Pr(X = x)}{Pr(X = x')} \); then, the relationship between LIP and DI follows the next lemma with proof provided in Appendix A.

**Lemma 2.** The relationship between LIP and DI is: \( \epsilon \)-LIP implies \((2\epsilon + D_X^\epsilon)\)-DI and \( \epsilon \)-DI implies \((\epsilon + D_X^\epsilon)\)-LIP.

So far, if a mechanism satisfies \( \epsilon \)-LIP, it implies \( \epsilon \)-MIP, \( \epsilon \)-MIL, \( 2\epsilon \)-LDP and \((2\epsilon + D_X^\epsilon)\)-DI. The main reasons why we choose to study LIP instead of other notions are as follows: (1) LIP is amenable to incorporate prior knowledge and design mechanisms compared to other context-aware definitions. (2) Compared with context-free LDP, LIP based mechanisms achieve much higher utility.

In the following sections, we address how LIP mechanisms can be designed according to the prior knowledge, and how such mechanisms improve the privacy-utility tradeoff for different types of applications.
3 MODELS AND PROBLEM FORMULATION

3.1 System and Threat Models

Consider a data aggregation system with $N$ users and a data curator. To answer query from the curator, each user locally generates data which is denoted as random variable $X_i$, taking value $x$ from the domain $D$ with the prior distribution of $\theta_X$, specifically, denote $P^i_x = Pr(X_i = x)$, where $i \in \{1, 2, ..., N\}$ is the user index. It is assumed that $X_i$s are independent of each other (and may have different distributions). To avoid potential privacy leakage, before publishing $X_i$, each user locally perturbs it by a privacy-preserving mechanism $M_i$. The output is denoted as $Y_i$ which takes value of $y$ from $D$ (we discuss the optimal support of $y$ in Sec. 4.1). The mechanism maps each possible input to each possible output with a certain probability (perturbation parameter). After receiving each perturbed data, the curator is allowed to further estimate and compute a statistical function of the collected data. The system is depicted in Fig. 3.

The curator is considered untrusted due to both internal and external threats. On one hand, users’ private data is profitable, and companies can be interested in user tracking or selling their data. On the other hand, data breaches happen from time to time due to hacking activities. The curator aims at performing accurate estimations using all the information above, but is also interested in inferring each user’s data value $X_i$. Denote the true aggregated result by $S = f(X)$, where $X = \{X_1, X_2, ..., X_N\}$. For different applications of data aggregation, the definition of $f(\cdot)$ varies. In this paper, two applications are considered:

- Weighted summation: the curator is interested in finding the summation over users’ data: $S = \sum_{i=1}^{N} (c_iX_i + b_i)$. When each $c_i = 1$ and $b_i = 0$, the application is equivalent to a direct summation, which is useful to find the average value;
- Histogram: the curator is interested in estimating how many people possess each of the data category in $D$, or classifying according to users’ data value: $S$ is a set of “categorized” data: $\{S_1, S_2, ..., S_d\}$, such that, $\forall k \in D$, $S_k = \sum_{i=1}^{N} I\{X_i = k\}$, where $I\{a=b\}$ is an indicator function, which is 1 if $a = b = 0$ if $a \neq b$. When there are only two categories, such application is also known as frequency estimation [17].

The curator (adversary) observes all the users’ outputs $Y = \{Y_1, Y_2, ..., Y_N\}$ and tries to obtain an estimation of $S$ using estimator $\hat{S}$.

In terms of prior availability, multiple scenarios could arise in practice which involve different assumptions of prior availability. For example, both the user and the curator know $\theta_X$ exactly, or one party is uncertain about $\theta_X$, or they possess asymmetric prior knowledge that can also be inaccurate. Within the scope of this paper, we assume that the curator always knows the exact $\theta_X$, and the algorithms/perturbation mechanisms that users deployed to publish their data. In the basic setting, we assume each user also possess the exact prior (same as the curator). Then we relax it and consider uncertain prior at the user side ($\theta_X^i$ is a subset of $P_X$). All the related symbols are listed in Table 1.

3.2 General Privacy and Utility Definitions

The privacy of each user’s is guaranteed by LIP and is parameterized by the privacy budget ($\epsilon$) in Definition 6. The smaller $\epsilon$ is, the stronger privacy guarantees the mechanism provides. Note that, for simplicity, we consider $\epsilon$ to be the same for all the users; however, it is straightforward to extend our model and results to the scenarios where different users are provided by different $\epsilon$s. When the exact prior $\theta_X^i$ is not available for each user, he/she defines $P_X^i$ to be the set of plausible priors including $\theta_X^i$, note that users are always allowed to enlarge the size of the $P_X^i$ to include $\theta_X^i$. For simplicity, we remove the conditional term in the probability metric, i.e., $Pr(X_i = x|\theta_X^i) = Pr(X_i = x)$, thus, the privacy constraints can be formulated as: $\forall i \in \{1, 2, ..., N\}, x, y \in D$, $\theta_X^i \in P_X^i$ there is

$$e^{-\epsilon} \leq \frac{Pr(X_i = x|Y_i = y)}{Pr(X_i = x)} \leq e^\epsilon.$$  \hspace{1cm} (8)

Denote the perturbation parameter: $q_{xy}^i \triangleq \{Pr(Y_i = y|X_i = x), \forall x, y \in D\}$, by Bayes rule, Eq. (8) can be expressed as:

$$e^{-\epsilon} \leq \frac{q_{xy}^i}{\sum_{x \in D} q_{xy}^i P^i_x} \leq e^\epsilon.$$  \hspace{1cm} (9)

Denote $q^i$ as the set of perturbation probabilities in $M_i$, then, when $\epsilon$ and each $P^i_x$ are given, the set of inequalities in Eq. (9) forms a feasible region $T^i$ for $q^i$, $\forall i \in \{1, 2, ...N\}$.

The definition of utility depends on the application scenario. For example, in statistical aggregation, the estimation accuracy is often measured by absolute error or mean square error [41][42]; in location tracking, it is typically measured by Euclidean distance [22]; under information theoretical framework, distortion is typically applied [29]. Without loss of generality, denote $U(S, \hat{S})$ as the utility.
In general, there is a tradeoff between utility and privacy. We can formulate the following optimization problem to find the optimal mechanism that yields the optimal tradeoff:

$$\begin{align*}
\max U(S, \hat{S}), \\
\text{s.t. } q^i \in T_i, \quad \forall i \in \{1, 2, \ldots, N\}.
\end{align*}$$ (10)

### 3.3 Problem Formulation

Focusing on the two applications discussed above, we define utility as the inverse of the Mean Square Error (MSE), which is also adopted in many other works on frequency/histogram estimation [17], [42], [43]. $U(S, \hat{S}) = -E(S - \hat{S})^2$, where $E(S, \hat{S}) = E[(S - \hat{S})^2]$. Note that, for weighted summation, the utility is data alphabet dependent while for histogram estimation, it is data alphabet independent, we show how MSE addresses these two different utilities in Sec. 4.4. Notice that, given $P^i$, the MSE $E(S, \hat{S})$ depends only on each user’s perturbation parameters: $\{q^i\}_{i=1}^N$, as any estimation $\hat{S}$ depends on the output $Y$ whose distribution is a function of $\{q^i\}_{i=1}^N$. Thus, maximizing the utility is equivalent to finding optimal parameters to minimize the MSE. As a result, the problem defined in Eq. (10) can be specified as:

$$\begin{align*}
\min \mathcal{E}(q^1, \ldots, q^N), \\
\text{s.t. } q^i \in T_i, \quad \forall i \in \{1, 2, \ldots, N\}.
\end{align*}$$ (11)

Denote this model of optimization problem as the Opt-LIP. From [44], it is well-known that the optimal estimator that results in the minimized mean square error (MMSE) is $\hat{S} = g(Y) = E[S|Y]$. Since $E[E[S|Y]] = E[S]$, $\hat{S}$ is an unbiased estimator. Therefore, we use the MMSE estimator in Eq. (11).

**Problem Decomposition:** Next, we show the problem defined in Eq. (11) can be decomposed into each local user.

Since we assume that each user’s input is independent of each other, all the $f(\cdot)$ functions above can be decomposed into local functions of each $X_i$, without loss of generality, we denote the local function for user $i$ as $f_i$. Thus we have local function in weighted summation as $f_i(X_i) = c_iX_i + b_i$; local functions in histogram are indicator functions (or vector): $\{j_i^k(X_i) = 1_{(X_i = k)}\}_{k=1}^B$.

In the local setting, users independently perturb their data, thus each of them results in an MSE in aggregation, which is denoted by $\tilde{E}_i = E[(f_i(X_i) - E[f_i(X_i)])^2]$. For the application of histogram, denote $\tilde{E}_i^k = E[(f_i^k(X_i) - E[f_i^k(X_i)])^2]$ as the MSE of aggregating the $k$-th data in the $i$-th user’s $X_i$, and the utility defined in Eq. (11) satisfies decomposition theorem with proof provided in Appendix B.

**Theorem 1.** The global optimization problem defined in Eq. (11) can be decomposed into $N$ local optimization problems, under independent user inputs.

$$\begin{align*}
\min_{\{q^i\} \in T_i} \mathcal{E}(q^1, \ldots, q^N) = \sum_{i=1}^N \min_{q^i \in T_i} \mathcal{E}_i(q^i).
\end{align*}$$ (12)

By Theorem 1 when the perturbation parameters of each user are optimal, the overall MSE of the mechanism achieves its minimum. In addition, each user can perform its local optimization independent of each other, which well suits the local setting. Next, we formulate the utility and privacy tradeoff as an optimization problem in each local case with an MSE of:

$$\begin{align*}
\tilde{E}_i(q^i) &= E[(f_i(X_i) - E[f_i(X_i)])^2] \\
&= E[E[(f_i(X_i) - E[f_i(X_i)])^2|Y_i]] \\
&= E[Var(f_i(X_i)|Y_i)]
\end{align*}$$ (13)

where $(a)$ follows the law of total variance.

Thus the MSE is a function of the variance of each user’s estimator. Define $\bar{X}_i = E[X_i|Y_i]$ as the local estimator for the $i$-th user, and we have $\bar{S} = \sum_{i=1}^N \bar{X}_i$, which follows the user independence assumption. Thus, local optimization problem for the $i$-th user can be reformulated as:

$$\begin{align*}
\min_{q^i} \tilde{E}_i(q^i) = \max \Var(\bar{X}_i), \\
\text{s.t. } q^i \in T_i, \quad \text{s.t. } q^i \in T_i.
\end{align*}$$ (14)

Which means, the optimal solutions are at the maximum of the variance of the estimator, subject to the LIP constraints.

### 4 Optimal Mechanism Design under LIP

In this Section, we study the privacy-utility tradeoffs according to different scenarios of $P^i_{X_i}$ including fixed prior, bounded priors, then, we compare the utility-privacy tradeoff of LIP to that of LDP (worst-case prior). Finally, we discuss how to realize each of the two applications with proposed frameworks.

#### 4.1 Optimal Perturbation Parameters for a Fixed Prior

Next, we derive the closed-form optimal solutions under the Opt-LIP model for any arbitrary but fixed prior. Under a randomized response perturbation mechanism, the perturbation channel and corresponding parameters are shown in Fig. 4. Denote $D = \{a_1, a_2, \ldots, a_d\}$, the prior distribution as $Pr(X_i = a_m) = P^i_{a_m}$, and the marginal distribution of $Pr(Y_i = a_k)$, the privacy metric becomes, $\forall m, k \in \{1, 2, \ldots, d\}$:

$$e^{-\epsilon} \leq \frac{Pr(X_i = a_m)}{Pr(X_i = a_m|Y_i = a_k)} \leq e^\epsilon.$$ (16)

By Bayes rules, Eq. (16) can be expressed as:
In the utility function of Eq. (14), \( \text{Var}[X_i] = \sum_{m=1}^{d} a_m^2 P_m - \sum_{m=1}^{d} a_m P_m^2 \), and the local estimator under the MIMO model: \( \hat{X}_i^m \) (where \( m \) stands for MIMO model) becomes:

\[
\hat{X}_i^m = E[X_i | Y_i] = \sum_{m=1}^{d} a_m \Pr(X_i = a_m | Y_i)
\]

\[
= \sum_{m=1}^{d} a_m \Pr(X_i = a_m | Y_i = a_k) \mathbb{1}_k^i,
\]

where \( \mathbb{1}_k^i \) is the indicator function of \( \mathbb{1}_k^i (Y_i = a_k) \), which is 1 if \( Y_i = a_k \) and 0 if not, thus \( \mathbb{1}_k^i \) can be regarded as a binary random variable which with the distribution of: 
\[
Pr(\mathbb{1}_k^i = 1) = \lambda_k^i \text{ and } Pr(\mathbb{1}_k^i = 0) = 1 - \lambda_k^i,
\]

As a result:

\[
\text{Var}(\hat{X}_i^m) = \sum_{m=1}^{d} \sum_{n=1}^{d} \sum_{k=1}^{d} a_m a_n q_{mk} q_{nk} \text{Var}[\mathbb{1}_k^i]
\]

\[
+ \sum_{m=1}^{d} \sum_{n=1}^{d} \sum_{k=1}^{d} a_m a_n q_{mk} q_{nk} \text{Cov}[\mathbb{1}_k^i, \mathbb{1}_l^i]
\]

\[
= \sum_{m=1}^{d} \sum_{n=1}^{d} \sum_{k=1}^{d} a_m a_n P_m P_n q_{mk} q_{nk} \left( \frac{q_{nk}^i (1 - \lambda_k^i)}{\lambda_k^i} - \sum_{l=1; l \neq k}^{d} q_{nl}^i \right)
\]

\[
= \sum_{m=1}^{d} \sum_{n=1}^{d} \sum_{k=1}^{d} a_m a_n P_m P_n q_{mk} q_{nk} \left( \frac{q_{nk}^i}{\lambda_k^i} - 1 \right)
\]

\[
= \sum_{m=1}^{d} \sum_{n=1}^{d} a_m a_n P_m P_n \left( \sum_{k=1}^{d} \frac{q_{mk} q_{nk}^i}{\lambda_k^i} - 1 \right).
\]

So far, Eq. (15) can be further expressed as \( \forall m, k \in \{1, 2, ..., d\} \):

\[
\text{max Var}(\hat{X}_i^m),
\]

\[
\text{s.t. } e^{-\epsilon} \leq \lambda_k^i q_{mk} \leq e^\epsilon.
\]

The global optimal solutions follow the next Theorem, with detailed proof provided in Appendix C.

**Theorem 2.** For the constrained optimization problem defined in Eq. (20), the optimal solutions for the \( i \)-th user are: \( q_{mm}^i = 1 - (1 - P_m^i)/e^\epsilon \), \( q_{mk}^i = P_k^i / e^\epsilon \), \( \forall m, k \in \{1, 2, ..., d\}, m \neq k \).

The constrained optimization problem defined in Eq. (20) can be visualized in Fig. 5 (taking a binary example), where the curves in the figure stand for the contour of \( \text{Var}(\hat{X}_i^m) \) and the shaded area stands for the feasible region of \( \mathcal{T}^i \) for a fixed prior and \( \epsilon \). Then the optimal solutions are found at the boundary of the feasible region, which are intersections of linear equations.

From Theorem 2 when \( \epsilon \) increases, \( \forall m \in \{1, 2, ..., d\} \), all the \( q_{mm}^i \)s are increasing while all the \( q_{ik}^i \)s are decreasing \( (m \neq k) \), and the value of \( q_{ik}^i \)s are proportional to \( P_k^i \)s, i.e., the optimal mechanism is more likely to output the values with larger priors.

For example, suppose \( \mathcal{D} = \{1, 2, 3\} \), for the \( i \)-th user: \( P_1 = 0.1, P_2 = 0.2, P_3 = 0.7 \). By Theorem 2, \( q_{11}^i = 1 - 0.9/e^\epsilon \), \( q_{22}^i = 1 - 0.8/e^\epsilon \), \( q_{33}^i = 1 - 0.3/e^\epsilon \), \( q_{21}^i = q_{31}^i = 0.1/e^\epsilon \).
### 4.2 Optimal Perturbation Parameters for Bounded Priors

Next, we consider the case where each $P^i_X$ includes multiple possible priors for the $i$-th user, in other words, each user has uncertainty on the exact prior. On the other hand, under the context-aware setting, it is assumed that the curator/adversary possesses the exact prior distribution. Such scenarios exist when users possess less information about $X$ than the curator. For example, the curator has recorded a full history of users’ previous released data in the server such that the curator is able to infer each user’s prior; another example is the curator can estimate a global prior for all the users by observing each user’s released data; the third example might be, the user is highly correlated with someone (such as family members or close friends) whose data has been collected or compromised, then the user’s prior can be inferred by the curator via the correlations.

In this uncertain prior model, the exact prior $\theta^i_X$ is not available for each user, as a result, the prior-dependent utility function defined in Eq. (14) cannot be calculable either. In such case, for each user, the local MSE function for him/her is determined by both of his/her perturbation parameters as well as exact prior distribution, i.e., $E_i(q^i)$ in Eq. (14) becomes $E_i(\theta^i_X, q^i)$.

As the $E_i(\theta^i_X, q^i)$ depends on the perturbation parameters $q^i$ as well as the exact prior distribution $\theta^i_X$. Under the uncertain prior model, One feasible minimax strategy for each user is to find the maximized $E_i(\theta^i_X, q^i)$ achieved by a prior of $\theta^i_X \in P^i_X$ and find $q^i$ which minimizes $E_i(q^i|\theta^i_X)$ (other feasible strategies exist, such as minimizing $E[E_i(\theta^i_X, q^i)]$, however are out of the scope of this paper). Thus the problem for the $i$-th user under the model with uncertain prior becomes:

$$\min_{q^i} \max_{\theta^i_X \in P^i_X} E_i(\theta^i_X, q^i),$$  \hspace{1cm} \text{s.t.} \quad q^i \in T_i. \hspace{1cm} \tag{21}$$

Note that the feasible region $T_i$ in Eq. (21) is different from the one defined for a fixed prior. It uses BP-LIP’s definition, i.e., LIP must be satisfied for a family of priors.

Denote this model with optimization problem in Eq. (21) as Opt-BP-LIP. Note that, the utility function in Eq. (21): $E_i(\theta^i_X, q^i) = \text{Var}(X_i) - \text{Var}(X^{bp}_i)$, where $X^{bp}_i$ is the optimal estimator at the curator. As $\text{Var}(X_i)$ depends only on the exact prior of $\theta^i_X$, the goal of each user is still to maximize $\text{Var}(X^{bp}_i)$. Thus Eq. (21) can be further expressed as:

$$\max_{q^i \in T^i} \min_{\theta^i_X \in P^i_X} \text{Var}(X^{bp}_i(\theta^i_X, q^i)). \hspace{1cm} \tag{22}$$

The optimal solutions of the MIMO model of each user with bounded priors depend on the concrete $P^i_X$ and thus can only be derived numerically (the comparison result is shown in Sec. 6). We next derive the optimal parameters under a binary model. Firstly, specify $P^i_X$ as $P^i_1 = \text{Pr}(X_i = 1) \in [a, b]$, where $0 \leq a \leq b \leq 1$. The optimal solutions correspond to the following proposition:

**Proposition 1.** For the constrained optimization problem defined in Eq. (22) with binary input/output, the optimal solutions for the $i$-th user are: $q^{i*}_{01} = \frac{1-a}{1-a+ce}$ and $q^{i*}_{10} = \frac{1-b}{1-a+ce}$.

---

### 4.3 Comparison with LDP

The original LDP notion is defined in the context-free setting, which is not directly comparable with LIP. In this part, we derive utility-privacy tradeoff for LDP notion in a context-aware setting: we assume the curator still possesses the prior distribution of each user’s local data, thus, the estimator is prior-dependent and the form is identical to $X^{m}_i$. As a result, the optimization problem of the LDP model has the same utility function as in Opt-LIP model, but subject to LDP constraints, i.e., the local optimal parameters for the $i$-th user correspond to the optimization problem:

$$\min_{q^i} E_i(q^i),$$  \hspace{1cm} \text{s.t.} \quad \text{Pr}(Y_i = y|X_i = x) \leq e^i, \forall y, x, x' \in \mathcal{D}. \hspace{1cm} \tag{23}$$

Denote this model with optimization problem in Eq. (23) as Opt-LDP. The privacy constraints in Eq. (23) can be expressed as: $\forall m, n, k \in \{1, 2, ..., d\}, m \neq n; \frac{q^{mk}_{nk} - q^{nk}_{mk}}{q^{mk}_{nk}} \leq e$. Learn from the proof of Theorem 2 in order to increase utility, the optimal mechanism increases $q^{i*}_{nk}$s while decreasing $q^{i*}_{mk}$s. Thus, the optimal solutions are at the boundaries of the LDP constraints of $\frac{q^{i*}_{nk}}{q^{i*}_{mk}} = e^i$. The solutions correspond to the following proposition:

**Proposition 2.** The optimal solutions of the Opt-LDP model are $q^{i*}_{min} = \frac{e}{e^i+1-e^i}$ and $q^{i*}_{max} = \frac{1}{e^i+1-e^i}$, $\forall m, k \in \{1, 2, ..., d\}, m \neq k$.

Similar results can also observed from [17], in which a different utility goal is defined.
Denote $E_{LIP}^i$ as the local MSE from collecting the $i$-th user’s data under LIP constraints and $E_{LDP}^i$ as that under LDP constraints. Comparing $E_{LIP}^i$ with $E_{LDP}^i$, we have the following proposition:

**Proposition 3.** Given an arbitrary but fixed prior distribution, $\forall \epsilon \in \mathbb{R}^+$, there is $E_{LIP}^i \leq E_{LDP}^i$.

**Proof.** Since $E_{LIP}^i$ and $E_{LDP}^i$ are the objective function evaluated at different optimal solutions satisfying corresponding privacy constraints. It suffices to show that the optimal perturbation parameters of LDP are within the feasible region of LIP. As $c$-LDP implies $c$-LIP, $\forall \epsilon \geq 0$, which means all the $q$’s that satisfying LDP automatically satisfies LIP.

Notice that, even the curator may take advantage of his prior knowledge to make a further estimation, LDP based mechanism still suffers a decreased utility than LIP because LIP also utilizes the prior knowledge for mechanism design. Also note that the Opt-LDP model and Opt-LIP model only differ in the feasible regions formed by corresponding privacy constraints, while the feasible region of LDP is fixed for all possible priors, the feasible region of LIP reshapes when the prior changes. Consider a worst-case scenario, and comparing the feasible regions formed by LIP and LDP constraints, we have the following remark:

**Remark 2.** The feasible region of LDP is the intersection of the feasible regions of LIP with all possible priors.

We visualize this remark combined with BP-LIP in Fig. 3 using a binary example, where we can observe the feasible regions of LIP with different priors ($P_i^q = 0.01, 0.25, 0.5, 0.75, 0.99$ respectively), the feasible region of LIP with a bounded set of priors ($P_i^q \in [0.5, 1]$) as well as the feasible region of LDP. To achieve high utility, $q_0^i$ and $q_1^i$ should be as small as possible. From Theorem 1 and Proposition 2, the optimal parameters for $q_0^i (q_1^i)$ under the three models are $(0.01, 0.99), (0.25, 0.75), (0.5, 0.5), (0.75, 0.25), (0.99, 0.01)$ for LIP with different priors; $(1, -1)$ for BP-LIP and $(1, -1)$ for LDP, respectively. Obviously, the optimal solution of LDP provides reduced utility compared to that of LIP and BP-LIP. The feasible region of BP-LIP is the intersection of the feasible regions of LIP with $P_i^q = 0.5$ and $P_i^q = 1$; The feasible region of LDP is the intersection of the feasible regions of LIP with all possible priors.

Next, we use a binary example to illustrate how the concrete prior distribution affects the utility resulted by LIP and LDP in terms of MSE.

When $X_i$ is binary, Opt-LIP results in local MSE of:

$$E_{LIP}^i = P_i^q (1 - P_i^q) (2e^{-\epsilon} - e^{-2\epsilon}),$$

in contrast to

$$E_{LDP}^i = P_i^q (1 - P_i^q) - \frac{[P_i^q (1 - P_i^q) (1 - e^{-\epsilon})]^2}{(1 + P_i^q + P_i^q e^{-\epsilon}) (e^{-\epsilon} - P_i^q e^{-\epsilon} + P_i^q)},$$

provided by Opt-LDP.

It is readily shown that $E_{LIP}^i \leq E_{LDP}^i$, where $E_{LIP}^i = E_{LDP}^i$ if $\epsilon = 0$ or $\infty$. We then take derivative with respect to $P^q_i$ over $\Delta E_i = E_{LDP}^i - E_{LIP}^i$, result shows that $\frac{\partial E_i}{\partial P_i^q} = 0$ when $P_i^q = 0.5$. Result also shows that when $|P_i^q - 0.5|$ increases, $\Delta E$ also increases.

### 4.4 Real-world Applications of LIP

Next, we discuss how to apply the model described above for the two applications.

#### 4.4.1 Application to (Weighted) Summation

Summation results usually measure an average property of the surveyed individuals, and it is straightforward to extend the direct summation to a weighted summation (plus an offset) for more general applications. For example, assume that the curator is interested in some particular users more than others, such as employer v.s. employees; adults v.s. children; the professionals v.s. amateurs. These “important” users’ data are assigned with larger coefficients than others. The offset can be used as a correction to the raw data.

For data summation, the collected data from each user is from a large domain. In this case, $f_i(X_i) = X_i$, and the MIMO model is suitable for this application. For weighted summation, $f_i(X_i) = c_iX_i + b_i$ and the aggregated result is $\hat{S}_{sum} = \sum_{i=1}^{N}(c_iX_i + b_i)$ with the estimator of $\hat{S}_{sum} = \sum_{i=1}^{N}(c_i\hat{X}_i + b_i)$, when $c_i$ and $b_i$ are known, the MSE becomes:

$$E[(\hat{S}_{sum} - \hat{S}_{sum})^2] = \sum_{i=1}^{N} \left( \sum_{i=1}^{N} (c_i(X_i - \hat{X}_i)) \right)^2$$

$$= \sum_{i=1}^{N} \left( \sum_{i=1}^{N} c_i^2(X_i - \hat{X}_i)^2 + \sum_{i,j=1}^{N} c_i c_j (X_i - \hat{X}_i)(X_j - \hat{X}_j) \right)$$

$$\leq \sum_{i=1}^{N} \sum_{j=1}^{N} c_i^2 E[(X_i - \hat{X}_i)^2],$$

where (a) follows the independent user assumption. By decomposition theorem, this problem becomes identical to that of the Opt-LIP model. Note that, when users have uncertain priors, as long as the curator possesses each accurate $\theta^q_X$, he is able to design each local unbiased estimator accordingly, which makes the global utility of $E[(\hat{S}_{sum} - \hat{S}_{sum})^2]$ decomposable.

**Remark 3.** The optimal perturbation parameters for the application of (weighted) summation follow the solutions of the Opt-LIP model in Theorem 1 while each local estimator changes from $X_i^m$ to $X_i^s = c_i X_i^m + b_i$. Similar results can be concluded in the uncertain prior models of Opt-BP-LIP and Opt-LDP with corresponding optimal solutions.

#### 4.4.2 Application to Histogram Estimation

Histogram is useful to estimate or compare the popularity or frequency of some categories. The difference between the application of the histogram and other applications is: even though for each user, the perturbation mechanism still takes one input data and outputs one data, the estimator is a random vector rather than a random variable, as the value of each data stands for a category. It can be viewed as a general form of survey, where a binary category is aggregated.
We can obtain the estimator of the histogram vector, $\hat{S}_{\text{hist}} = \{\hat{S}_1, \hat{S}_2, ..., \hat{S}_d\} = \{E[S_1|Y], E[S_2|Y], ..., E[S_d|Y]\}$, with each entry $E[S_m|Y]$:

$$E \left\{ \sum_{i=1}^{N} \mathbb{1}_{\{X_i = a_m\}} | Y \right\} = \sum_{i=1}^{N} Pr(X_i = a_m | Y_i). \quad (26)$$

Thus the mean square error of the estimation is

$$\sum_{m=1}^{d} E \left\{ \left[ \sum_{i=1}^{N} (\mathbb{1}_{\{X_i = a_m\}} - E[\mathbb{1}_{\{X_i = a_m\}} | Y_i]) \right]^2 \right\} \quad (27)$$

$$= \sum_{m=1}^{d} \sum_{i=1}^{N} E[(\mathbb{1}_{\{X_i = a_m\}} - E[\mathbb{1}_{\{X_i = a_m\}} | Y_i])^2]$$

$$= \sum_{i=1}^{N} \sum_{m=1}^{d} \left\{ \text{Var}(\mathbb{1}_{\{X_i = a_m\}}) - \text{Var}(E[\mathbb{1}_{\{X_i = a_m\}} | Y_i]) \right\}.$$ 

The (a) of Eq. (27) is because each user’s local error is independent, and the expectation of the unbiased estimator is identical to that of the estimated value. Hence, the problem of estimating a histogram can be formulated as, $\forall m, k \in \{1, 2, ..., d\}, i \in \{1, 2, ..., N\}$:

$$\min \text{Eq. (27)},$$

s.t. $e^{-\epsilon} \leq \frac{\lambda_i}{q_{mk}} \leq e^{\epsilon}. \quad (28)$$

The optimal solution for the above problem are presented in the following corollary:

**Corollary 1.** The optimal solutions of the problem defined in Eq. (28) follow Theorem 2 with each local estimator of $\hat{X}_i^p = \{Pr(X_i = a_1 | Y_i), Pr(X_i = a_2 | Y_i), ..., Pr(X_i = a_d | Y_i)\}$. Similar results can be concluded in the uncertain prior models of Opt-BP-LIP and Opt-LDP with corresponding optimal solutions.

Proof is provided in Appendix F.

Note that, by defining MSE as the utility, the function of (weighted) summation depends on the data alphabets, which means the difference of data values during aggregation also determines the performance of the mechanism. On the other hand, the utility function of histogram estimation is data alphabet independent, as the aggregated value in each class is the number of users. However, MSE adequately describes the utility in each scenario by designing local function $f_i()$ for each user. For summation, $f_i$ is value-dependent, while for histogram, $f_i$ is designed with indicator functions that transfer the data alphabets into a sequence of probabilities.

**5 Optimal Mechanism Design with Latent Variable for A Given Prior**

In many applications, the user’s secret information to be protected is different from but correlated with the data being collected. For example, wearable sensors or fitness trackers can collect data about a person’s acceleration, body temperature, heart rate and etc, to infer his/her daily activity pattern or exercise amounts; however he/she may wish to prevent the inference of sensitive disease information from those data, as such data may lead to discrimination from insurance companies if they obtain the data. As another example, smart meters collect people’s daily usage of electricity and use statistical information to optimize the operation of the grid, but each individual household does not wish to reveal which appliance they are using at a given time. Similarly, big companies collect people’s location data for statistical analysis, but a user may want to hide some specific event/pattern in her location trace, or the private social relationship with another user.

To cope with the above scenarios, we extend the definition of LIP to consider latent variables.

**5.1 Utility-Privacy Formulations with Latent Variable**

The model is depicted in Fig. 7. For each user, the raw data is denoted as $X_i$, which is correlated with a latent variable $G_i$ with some joint distribution. Denote $G$ as the universe of all the latent variables, and $T_{g,x}$ as the conditional probability of $Pr(X_i = x | G_i = g)$ for all $x \in \mathbb{D}$ and $g \in G$.

Under LIP, the privacy constraints can be formulated as: $\forall i \in \{1, 2, ..., N\}, \forall g \in G$ and $y \in \mathbb{D}$, there is

$$e^{-\epsilon} \leq \frac{Pr(G_i = g | Y_i = y)}{Pr(G_i = g)} \leq e^{\epsilon}. \quad (29)$$

By Bayes rules, given the values of $T_{g,x}$s for all $x \in \mathbb{D}$ and $g \in G$, the metric in Eq. (29) can be further expressed as a function of $q_{xg}$:

$$e^{-\epsilon} \leq \frac{\sum_{x \in \mathbb{D}} q_{xg} T_{g,x}}{\sum_{x \in \mathbb{D}} q_{xg} P_{x}} \leq e^{\epsilon}. \quad (30)$$

The constrained optimization problem in Eq. (15) under the model with latent variables can be formulated as, $\forall i \in \{1, 2, ..., N\}, g \in G$, $y \in \mathbb{D}$:

$$\max \text{Var}(\hat{X}_i^{(m)}),$$

s.t. Eq. (30).

In general, the closed-form optimal solution for the constrained optimization problem of Eq. (31) cannot be directly derived, as the number of linear constraints is quadratically proportional to the dimensions of $X_i$ and $G_i$, and the valid constraints depend on the concrete values of the priors and correlations. We numerically present the results and show the properties of the model with latent variables in Sec. 6.

**5.2 Optimal Solutions under Binary Model**

Next, we derive closed-form optimal solutions for the model with binary input/output, which is arbitrarily correlated with a binary latent variable. The binary model is...
widely used for survey, where each individual’s data is
first mapped to one bit, then randomly perturbed before
publishing to the curator.

In this model, \(\mathcal{D} = \mathcal{G} = \{0, 1\}\) (shown in Fig. 8). Var(\(X_i\)) in Eq. (14) becomes \(\mathcal{P}_i^* (1 - \mathcal{P}_i^*).\) Denote the parameters as:

\[
\begin{align*}
Pr(Y_i = 1|X_i = 0) &= q_{i0}^i, \\
Pr(Y_i = 0|X_i = 1) &= q_{i1}^i.
\end{align*}
\]

Thus, the local MMSE estimator \(\hat{X}_i^h\) becomes:

\[
\hat{X}_i^h = E[X_i|Y_i] = \mathcal{P}_i^* \left[ q_{i0}^i \lambda_0^i (1 - Y_i) + \frac{1 - q_{i0}^i}{\lambda_0^i} Y_i \right],
\]

where \(\lambda_0^i = Pr(Y_i = 0) = (1 - P_1) (1 - q_{i0}^i) + P_1 q_{i1}^i\) and \(\lambda_1^i = Pr(Y_i = 1) = (1 - P_1) q_{i0}^i + P_1 (1 - q_{i1}^i).\) Then,

\[
\text{Var}(\hat{X}_i^h) = \text{Var} \left( \mathcal{P}_i^* \left[ q_{i0}^i \lambda_0^i (1 - Y_i) + \frac{1 - q_{i0}^i}{\lambda_0^i} Y_i \right] \right) = \frac{1 - q_{i0}^i - q_{i1}^i \lambda_0^i}{\lambda_0^i \lambda_1^i}.
\]

For the privacy constraints, by Eq. (29), when the perturbation mechanism satisfies \(\epsilon\)-LIP, the feasible region \(\mathcal{T}_i^h\) is formed by:

\[
e^{-\epsilon} \leq \{F_1, F_2\} \leq e^{\epsilon}, \quad \forall i = 1, 2, \ldots, N.
\]

where \(F_1, F_2\) are directly derived from Eq. (30):

\[
\begin{align*}
F_1(q_{i0}^i, q_{i1}^i, g) &= \frac{1 - q_{i0}^i T_{q0}^i + q_{i1}^i T_{q1}^i q_{i1}^i}{q_{i0}^i P_1^i (1 - q_{i0}^i) (1 - P_1^i)} + \frac{1 - q_{i0}^i}{\lambda_0^i \lambda_1^i}, \\
F_2(q_{i0}^i, q_{i1}^i, g) &= \frac{q_{i0}^i - q_{i1}^i T_{q0}^i + q_{i1}^i T_{q1}^i}{(1 - q_{i0}^i) P_1^i (1 - q_{i0}^i) (1 - P_1^i)}.
\end{align*}
\]

So far, the utility-privacy tradeoff of the binary model with latent variables can be formulated as:

\[
\max_{(q_{i0}^i, q_{i1}^i) \in \mathcal{T}_i^h} \text{Var}(\hat{X}_i^h). \tag{35}
\]

We next derive the optimal \(q_{i0}^i\) and \(q_{i1}^i\); firstly, define \(T_{q0}^i = \max_{g \in \mathcal{G}} T_{q0}^i \) and \(T_{q1}^i = \min_{g \in \mathcal{G}} T_{q1}^i.\) Then the optimal \(q_{i0}^i\) and \(q_{i1}^i\) correspond to the following theorem:

\[
\text{Theorem 3. The optimal } q_{i0}^* \text{ and } q_{i1}^* \text{ of the problem defined in Eq. (35) are:}
\]

\[
\begin{align*}
q_{i0}^* &= \max \{0, \frac{T_{q0}^i - P_1^i e^{\epsilon}}{(e^{\epsilon} + 1)(T_{q0}^i - P_1^i)}\} \frac{P_1^i - T_{q1}^i e^{\epsilon}}{(e^{\epsilon} + 1)(P_1^i - T_{q1}^i)} \\
q_{i1}^* &= \max \{0, 1 + T_{q0}^i e^{\epsilon} - e^{\epsilon} - P_1^i (1 + P_1^i e^{\epsilon} - e^{\epsilon} - T_{q1}^i)\}.
\end{align*}
\]

The proof is shown in Appendix E.

Compare with the optimal solutions in Theorem 2, we have that the optimal \(q_{i0}^*\) and \(q_{i1}^*\) in the model with latent variables are always smaller than that of the Opt-LIP model. Note that, when \(X_i = G_1, q_{i0}^* = P_1^i / e^{\epsilon}, q_{i1}^* = (1 - P_1^i) / e^{\epsilon},\) which means these two models are equivalent. On the other hand, if \(\max \{T_{q0}^i / P_1^i, T_{q1}^i / T_{q0}^i\} \leq e^{\epsilon}, q_{i0}^* = 0; \) If \(\max \{1 - T_{q0}^i / (1 - P_1^i), (1 - P_1^i) / (1 - T_{q1}^i)\} \leq e^{\epsilon}, q_{i1}^* = 0;\) when both conditions are satisfied, \(X_i\) is directly published.

Key insights from the binary model with latent variables is, when \(X\) is highly correlated with \(G\) (\(T_{q1}^i\) is large and \(T_{q0}^i\) is small), \(X\) should be heavily perturbed in order to protect \(G;\) When \(X\) is almost independent of \(G\) (\(T_{q0}^i\) and \(T_{q1}^i\) are close to \(P_1^i\)), \(X\) can be released with slight perturbation, such insight is depicted in Fig. 9.

5.3 Properties of MIMO model with latent variables

While the closed-form solutions for the MIMO model with latent variables are not straightforward, with the key insights of the binary model, we next derive some properties for the MIMO model with latent variables.

Firstly, the optimal solution in Theorem 2 is also a feasible solution of the problem defined in Eq. (31):

When \(Y_i\) is released satisfying \(\epsilon\)-LIP with respect to \(X_i,\) the privacy constraints in Eq. (30) is bounded by:

\[
e^{-\epsilon} \leq \sum_{x \in \mathcal{D}} q_{xy}^i T_{xy}^i \leq \sum_{x \in \mathcal{D}} q_{xy}^i P_{xy}^i \leq \max_{x \in \mathcal{D}} q_{xy}^i T_{xy}^i \leq e^{\epsilon}
\]

Which means Eq. (30) is satisfied.

As a result, protecting the privacy of a latent variable rather than the raw data enlarges the feasible region of the perturbation parameters, and hence, an increased utility can be achieved. We next show that, under some conditions, the privacy requirements are met without introducing noise.
Suppose for some \( \forall Y \) parable to the absolute error. Note that, doing so does not violate the privacy requirement. Such property of the model \( \varrho \) in Eq. (37) equals to \( \max \{ \frac{\max_{a \in G} T_{ga}}{P_a}, \frac{P_a}{\min_{a \in G} T_{ga}} \} \leq \epsilon^e \), \( \text{(37)} \) the optimal \( q_{ma} = 0 \) and \( q_{na} = 1 \), \( \forall m \neq a \).

\( \text{Proof.} \) Suppose for some \( a \in D \), \( q_{aa} = 1 \) and \( q_{ma} = 0 \), based on Eq. (20), \( \forall g \in G \), there is \( e^{-e} \leq \frac{T_{gma}}{P_a} \leq e^e \).

On the contrary, if this condition is satisfied, to maximize utility, the mechanism decreases \( q_{ma}^* \) while increasing \( q_{mm}^* \) thus \( q_{ma}^* = 0 \) and \( q_{aa}^* = 1 \).

Which means if \( X_i = a \), the mechanism directly releases \( Y_i = X_i \); It's straightforward to extend to: if \( \forall x \in D, \) \( \max \{ \frac{\max_{a \in G} T_{xa}}{P_a}, \frac{P_a}{\min_{a \in G} T_{xa}} \} \leq \epsilon^e \), then the mechanism directly releases \( Y_i = X_i \).

Notice that, \( \forall x \in D \) and \( \forall g \in G \), the bounded ratio in Eq. (27) equals to 1 when \( X_i \) and \( G_i \) are independent, which means directly releasing \( X_i \) leaks no information about \( G_i \), if the ratio is bounded close to 1, and the closeness is measured by \( e^{-e}, e^e \), directly releasing \( X_i \) also does not violate the privacy requirement. Such property of the model with latent variables can be better illustrated in Fig 10.

6 EVALUATION

In this Section, we present simulation results to validate our analytical results. In the first part, we validate our analysis using synthetic data and via Monte-Carlo simulation. With synthetic data, we show the advantages of the context-aware privacy notion (based on LIP) versus the context-free notion (based on LDP) by comparing their utility-privacy tradeoffs. Then we compare different models of LIP and LDP and show how the dataset domain affects the results. By Monte-Carlo simulation, we consider a latent variable \( G_i \) that is arbitrarily correlated with \( X_i \) and compare the utility-privacy tradeoffs provided by local models. In the second part, we evaluate with real-world datasets: Gowalla (location check-ins) and Census Income (People income survey).

We evaluate utility by square root average MSE, in order to normalize the influence of user count, also to make it comparable to the absolute error. Note that, doing so does not affect the optimalities in any of our optimization problems.

In addition, since LIP achieves a relaxed privacy level than LDP, it is difficult to compare their utilities under the same privacy guarantee. Thus, we compare their optimal utilities under any given privacy budget \( \epsilon \). All the simulations are done in Matlab (R2016a) on a Dell desktop (OptiPlex 7040; CPU: Intel (R) Core (TM) i5-6500 @ 3.2GHz; RAM 8.0 GB; OS: windows 64bit).

6.1 Simulation Results with Synthetic Data

6.1.1 Benefit of Context-awareness

Firstly, we would like to compare the utility-privacy tradeoffs between \( \epsilon \)-LIP and \( \epsilon \)-LDP using the binary model. The goal is to show the advantage of our proposed context-aware notion versus context-free notion of LDP. Intuitively, the utility gain of the former can be attributed to two factors: 1) using the prior in the MMSE estimator, which improves the accuracy compared with estimators that do not use prior knowledge; 2) the privacy guarantee of LIP is relaxed compared with LDP, by explicitly modeling prior in the definition. As a result, less noise is needed to satisfy the same privacy budget \( \epsilon \). To decouple the influence of the above two factors, we compare the utility-privacy tradeoff of LIP notions with two other LDP notions: for the first LDP notion, we use prior-independent estimator \( \hat{C} \) [17]. This model treats \( X_i \) as instances rather than random variables:

\[
\hat{C} = \frac{\sum_{i=1}^{N} Y_i - N p_i}{1 - 2p_i}, \quad \text{(38)}
\]

where \( q_{mk} = p_i = \frac{e^e}{e^e + d - 1}, \forall m \neq k \) is the optimal perturbation parameter. This is an unbiased estimator, which results in an MSE of:

\[
E[(S - \hat{C})^2] = \text{Var}[\hat{C}] = \frac{N(d - 2 + e^e)}{(e^e - 1)^2}. \quad \text{(39)}
\]

The second LDP model is the Opt-LDP model, which is discussed in Section 4.3.

The comparison is shown in Fig. 11 where \( \epsilon \) ranges from 1 to 5 with a step of 0.5. For now, we assume that \( N \) users share the same global prior. We can observe that, the square root average MSE of \( \epsilon \)-Opt-LIP is always smaller than that of “context-free LDP” under any given \( \epsilon \). When \( P_1 = 0.5 \) (prior is uniformly distributed), the distance between
these two models is smaller; when the prior is more skewed, advantage of the former is even enhanced. This validates the benefit of the context-aware estimator. On the other hand, by comparing the curves of \(\epsilon\)-Opt-LDP and \(\epsilon\)-LIP (using the same MMSE estimator), the error of Opt-LIP is always smaller than that of Opt-LDP, and the gap between the two models increases when \(P_1 = 0.99\). This result confirms that our relaxed prior-aware privacy notion leads to an enhanced utility. When \(P_1 = 0.99\), users’ inputs are highly certain, merely considering prior in the estimator can already result in accurate aggregation. Thus the advantage of the \(\epsilon\)-LIP is even enhanced.

### 6.1.2 Impact of Prior Uncertainty, Domain Size and Correlation with Latent Variable on Different models

We now evaluate the utility-privacy tradeoffs of the LIP notion when \(D\) has a large domain. The number of users and the domain size of each user are fixed as \(N = 5000\) and \(|D| = 5\), respectively. Without loss of generality, we assume \(D = \{0, 1, 2, 3, 4\}\), with the prior of each value randomly generated for 5 times. The set of priors are used as the bounded range (all plausible prior distributions), and one of them is used as the true prior. The utility-privacy tradeoffs are shown in Fig. 12(a). The figure shows that the \(\epsilon\)-Opt-LIP provides the most enhanced utility compared with other models; The \(\epsilon\)-LDP is sandwiched between \(\epsilon\)-LIP and \(2\epsilon\)-LIP; Moreover, the \(\epsilon\)-BP-LIP also provides better utility compared to \(\epsilon\)-LDP. According to the analysis in Sec. 4.3, when the bounded prior set includes all possible priors, \(\epsilon\)-WC-LIP is equivalent to \(\epsilon\)-LDP.

We next compare how the data domain affects the LIP and LDP: Consider 5000 users are in the system with the domain size of each user’s data from \(|D| = 2\) to \(|D| = 20\). We then fix \(\epsilon = 1\) and show the utilities with different input domain size. In Fig. 12(b), we observe that when \(|D|\) is small, the \(\epsilon\)-MIMO-LDP model provides better utility than the \(\epsilon/2\)-MIMO-LIP model. However, as \(|D|\) increases, the \(\epsilon/2\)-MIMO-LIP eventually outperforms \(\epsilon\)-MIMO-LDP. We can also observe that both the LDP and LIP models suffer from decreased utility when the \(|D|\) increases, but the LIP models decrease linearly while the LDP model decreases faster than that. In comparison, central IP suffers the least influence from the increasing size of the domain.

Then, we study the cases with a latent variable correlated with each \(X_i\). We use Monte-Carlo Simulation to study the convergence of performance when \(N\) increases. Fig. 12(c) shows the comparison among the three models described above. For the dataset, we assume each user’s data \(X\) has an arbitrary domain size between 2 and 10, the prior of the data is then randomly sampled according to the domain size. The dataset also contains a latent variable \(G\), which has the same data size with \(X\) and a randomly generated correlation with \(X\): \(TG_X\). In the local setting, we assume that each user publishes \(Y_i\) using the LIP/LDP models with perturbation parameters as solutions of the optimization problem defined in Eq. (15) (the problem is solved by the built-in optimizer of Matlab). The curator aggregates data using corresponding estimators discussed above. The error is measured by the averaged squared error, which is derived from 10000 times simulations, which are shown in Fig. 12(c). We can observe that \(\epsilon\)-LIP always provides higher utility than \(\epsilon\)-LDP under any \(\epsilon\). Also, the curve of \(\epsilon\)-LDP is almost sandwiched between the ones of \(\epsilon/2\)-LIP and \(\epsilon\)-LIP. Notice that the error diminishes to 0 as \(\epsilon\) increases, that is because the correlation between the raw data and the private latent variable is weak. For a given \(\epsilon\), directly releasing the raw data is still secure for the privacy of the latent variable.

### 6.2 Simulation with Real-world Datasets

#### 6.2.1 Location Check-In Dataset

In this Section, we compare the performance of different models with the real-world dataset Gowalla, which is a social networking application where users share their locations by checking-in. There are 6,442,892 users in this dataset. For each user, a trace of his/her check-in locations is recorded. For this dataset, we wish to estimate a histogram of users’ last check-in location. We first divide the area into 36 \times 36 districts, then map each user’s locations into districts. Each user’s past check-in locations are used for calculating a global prior of the last check-in location for all the users. As we studied in Section 4.4.2, for each user, the last check-in location is perturbed according to the LIP (LDP) channel and a random vector estimator is used for the curator to estimate the histogram. The results are shown in Fig. 13(a) where similar trends can be observed as in the empirical results. Note that compared with the theoretical results from Fig. 7, the advantage of LIP is even enhanced in Fig. 13(a) that is because the theoretical analysis uses data from a domain with \(|D| = 20\). On the other hand, in the dataset of Gowalla, the input data is from a domain with \(|D| = 36 \times 36\), even though many of the districts has 0 users checking-in, which results in zero priors for these districts. Based on the MIMO perturbation mechanism, for those districts with 0 prior, the system will also never output those districts, as a result, the data domain is equivalent to a much decreased one. Nevertheless, the effective domain size is \(|D| = 83\), which is much larger than 20. When domain size is larger, the advantage of LIP is enhanced.

#### 6.2.2 Latent Variable Privacy for Dataset of Annual Income Survey

Next, we testify our analysis of the model with latent variables by simulation on a real-world dataset: “Census income” (Adult dataset). This is a census survey dataset in which 48842 users’ personal information is listed, including 14 attributes, such as: age, work class, marriage, race, gender, education, and annual income. We assume each user’s data is published and collected independently. In the field of machine learning, the Adult dataset is usually used for predicting whether each user’s annual income is over 50k dollars by training on all the personal information (taken as features). In this experiment, we want to aggregate users’ work classes while protecting the annual incomes. In this dataset, the aggregated data \(X_i\) work class, has a domain size of 8: \{Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked\}. Each user’s annual income, \(G_i\), also has a domain size of 8: \{below 20k, 20k-30k, 30k-40k, 40k-50k, 50k-60k, 60k-70k, 70k-80k, over 80k\}. We use number 0 to 7 to stand for each of them and statistically calculate the frequency of each value to be the priors. We then, find the correlation
between each user’s work class and the annual income by deep learning (built-in network of Tensorflow). Then each user publishes his work class by the Opt-LIP/ Opt-LDP mechanism described in Section IV. The comparison is shown in Fig. 3(b). From which, we observe that the proposed Opt-ϵ-LIP model provides better utility than ϵ-LDP. Compared with Monte-carlo simulations, with this dataset, each model requires a larger ϵ to diminish to 0, because the latent variable G is highly correlated with the X. In order to protect the privacy of G, it requires larger probabilities to be perturbed than the cases with arbitrary correlations between X and G.

From the experiment results, we have the following insights: a) context-aware privacy notions achieve better utility than context-free notions, and when the prior is more skewed, the advantage becomes even enhanced; b) LIP notion achieves better utility than the LDP notions when using the same prior related estimator, the utility gain lies in measuring the prior knowledge in the privacy notion. c) When the data domain increases, the utility under each notion decreases, but the decreased amount of the LIP notion is smaller than that of the LDP notion; d) Utilities of the models with latent variables are higher than those without because the collected data becomes less sensitive. When the correlation between X and G is weak, for some ϵ, X can be directly published to achieve 0 MSE.

7 CONCLUSION

In this paper, the notion of local information privacy is proposed. As a context-aware privacy notion, it provides a relaxed privacy guarantee than LDP by introducing prior knowledge in the privacy definition while achieving increased utility. Combined with an MMSE estimator which also leverages prior knowledge, enhanced gains in utility can be obtained. We implement the proposed LIP notion into data aggregation framework and derive the utility privacy tradeoff, which is to minimize the MSE between the function of the collected data and the estimation, while protecting the privacy of the raw data or a private latent variable that is arbitrarily correlated with the collected data. We then theoretically compare the proposed LIP framework with those satisfying LDP definition. Finally, we use both synthetic and real-world data to show that, the proposed LIP framework provides better utility than LDP, especially when the prior information is more skewed and when the domain size is large.

APPENDIX A

PROOF OF LEMMA 1

Proof. When ϵ-LIP is satisfied, the privacy metric of DI can be expressed as:

\[
\frac{Pr(Y = y|X = x)}{Pr(Y = y|X = x')} \leq e^{\epsilon - D_N^X}. \]

Then we have:

\[
Pr(Y = y) = \sum_{x \in D} Pr(Y = y|X = x)Pr(X = x) \leq e^{\epsilon + D_N^X} Pr(Y = y|X = x')Pr(X = x) \leq e^{\epsilon + D_N^X} Pr(Y = y|X = x').
\]

Similarly, \( Pr(Y = y) \geq e^{-\epsilon - D_N^X} Pr(Y = y|X = x'). \) Thus \( (\epsilon + D_N^X)\)-LIP is satisfied.

APPENDIX B

PROOF OF THEOREM 1

Proof. The MMSE estimator \( \hat{S} \) can be expressed as:

\[
E[S|Y] = E[f(X)|Y] = E[f(X_1, X_2, ..., X_N)|Y]
\]

\[
(a) E[f_1(X_1)|Y] + E[f_2(X_2)|Y], ..., + E[f_N(X_N)|Y] \]

\[
(b) \sum_{i=1}^{N} E[f_i(X_i)|Y],
\]
where (a) in Eq. (40) is due to the independence of $X_i$s, and (b) is because $X_i$ is only correlated with $Y_i$ in the output sequence. Thus, $E(S, \tilde{S})$ can be derived as:

$$E(S, \tilde{S}) = E \left[ \left( \sum_{i=1}^{N} \left( f_i(X_i) - E[f_i(X_i)|Y_i] \right) \right)^2 \right]. \quad (41)$$

Note that, for the application of histogram, the error vector $\tilde{S}$ forms an error vector of $(\tilde{S}_k, \tilde{S}_k')_{k=1}^d$. By the definition of second order norm, the mean square error of this case is:

$$E(\tilde{S}_k, \tilde{S}_k') = \sum_{k=1}^{d} E \left[ \left( \sum_{i=1}^{N} \left( f_k(X_i) - E[f_k(X_i)|Y_i] \right) \right)^2 \right],$$

where $f_k(X_i) = 1 \{ x_i = k \}$.

We next show that in general, the total MSE can be decomposed into the summation of local MSEs (the proof of histogram is shown in Appendix E).

$$E(S, \tilde{S}) = E \left[ \left( \sum_{i=1}^{N} \left( f_i(X_i) - E[f_i(X_i)|Y_i] \right) \right)^2 \right] = \sum_{i=1}^{N} E \left[ f_i(X_i) - E[f_i(X_i)|Y_i]^2 \right]$$

$$- 2 \sum_{j=1,j \neq i}^{N} E \left[ (f_j(X_j) - E[f_j(X_j)|Y_j])(f_i(X_i) - E[f_i(X_i)|Y_i]) \right].$$

The cross terms are 0 because $\forall j,l \in \{1, ..., N\}$ and $j \neq l$: $E(\tilde{f}(f_j(X_j) - E[f_j(X_j)|Y_j])(f_i(X_i) - E[f_i(X_i)|Y_i])) = \ldots$

$$= E(\tilde{f}(f_j(X_j) - E[f_j(X_j)|Y_j])(f_i(X_i) - E[f_i(X_i)|Y_i]))$$

$$= E(\tilde{f}(f_j(X_j) - E[f_j(X_j)|Y_j])(f_j(X_j) - E[f_j(X_j)|Y_j])) \quad \text{where } E(\tilde{f}(f_j(X_j) - E[f_j(X_j)|Y_j])) = 0, \quad \text{because the estimators are unbiased. Thus, } E(S, \tilde{S}) = \sum_{i=1}^{N} \tilde{E}(q^i).$$

We next show that the global optimal solutions (perturbation parameters) satisfy each local privacy constraint:

Assume that for each user, the minimizer $\tilde{E}_i(q^i) = e_i$ is achieved at $q^i \in T_i$, then $E(q^1, ..., q^N) = \sum_{i=1}^{N} e_i$.

If for some user “k” who takes parameters $q^k \in T_k$, by assumption, we know that $\tilde{E}_k(q^k) \geq e_k$ Thus,

$$\sum_{i=1}^{N} \tilde{E}_i(q^i) + \tilde{E}_k(q^k) + \sum_{i=k+1}^{N} \tilde{E}_i(q^i) \geq \sum_{i=1}^{N} e_i.$$  

That means the minimal value of $E(q^1, ..., q^N)$, where $q^i \in T_i, \forall i \in \{1, ..., N\}$ can be achieved if for each user, $q^i = q^i = e_i$.

\[\square\]

APPENDIX C

PROOF OF THEOREM 2

Proof. Notice that $\text{Var}[X_i]$ is a non-negative constant, thus minimizing MSE is equivalent to maximize $\text{Var}[X_i]$.

Step 1. Regardless of the privacy constraints:

Minimized solution:
Consider a set of parameters: $q_{\text{min}}$, when $q_{nk} = \lambda_k$, $\forall n, k \in \{1, 2, 3, ..., d\}$, $\text{Var}[X_i] = 0$. Since $\text{Var}[X_i] \geq 0$, the solution of $q_{nk} = \lambda_k$ results in a minimal value of $\text{Var}[X_i]$.

Maximized solution: Consider a set of parameters: $q_{\text{max}}$, assume that for all $k = 1, 2, ..., d$, $q_{kk} = 1$ and $q_{kl} = 0$ for all $l \neq k$. Under this solution, $\lambda_k = P_k$ and

$$\sum_{n=1}^{d} \sum_{m=1}^{d} a_m a_n P_m P_n q_{nm} \left( \frac{q_{nk}}{\lambda_k} - 1 \right)$$

$$= \sum_{n=1}^{d} a_n^2 P_n^2 (1 - P_n) - \sum_{n=1}^{d} \sum_{m=1}^{d} a_m a_n P_m P_n = \text{Var}[X_i].$$

Notice that $\tilde{E}_i \geq 0$, $\text{Var}[X_i] \geq \text{Var}[X_i]$. Thus, the solution of $q_{kk} = 1$ and $q_{kl} = 0$, $\forall k = 1, 2, ..., d, l \neq k$ results in the maximum value of $\text{Var}[X_i]$.

Next, investigate the monotocity of the region between minimum and maximum:

Taking derivative with respect to $q_{lk}$, becomes

$$\frac{\partial \text{Var}[X_i]}{\partial q_{lk}}$$

$$= \frac{1}{\lambda_k^2} a_l \lambda_k \left( 2 \sum_{m=1}^{d} (a_m q_{mk} - a_j \lambda_k) \right) - P_l \left( \sum_{m=1}^{d} a_m q_{nm} \right)^2$$

$$= \frac{a_l q_{lk}}{\lambda_k} \left( \sum_{m \neq l} a_m q_{mk} (1 - P_k) q_{lk} - \lambda_k \right).$$

From Eq. (43), we can observe that the station point of $q_{lk}$ is $\lambda_k$, which we know is the minimal value and $\text{Var}[X_i]$ is monotonically increasing when $q_{lk} > \lambda_k$; $\text{Var}[X_i]$ is monotonically decreasing when $q_{lk} < \lambda_k$. As a result, without considering the privacy constraints, the optimal solutions of each $q_{mn}$ is either 0 or 1. We next show that the maximum value of $\text{Var}[X_i]$ can only be achieved by the solutions discussed above.

Now, assume that for the data value $l$, there is a subset of index $S$ s.t. $q_{lk} \neq 1 \neq 0$, for any $k \in S$. Denote $X$ as the estimator using $q_{\text{max}}$ and $\tilde{X}$ as the estimator using $q_{\text{min}}$ but the parameters for data value $l$ are substituted according to the subset. Regardless of the constraints, compare with the variance of $\text{Var}[X_i]$ and $\text{Var}[\tilde{X}_i]$, we have:

$$\text{Var}[\tilde{X}_i] - \text{Var}[X_i]$$

$$= \sum_{k=1}^{d} a_k^2 P_l + \frac{P_i}{2} + \frac{P_k}{2} + \frac{P_k}{2} + \frac{P_i}{2}$$

$$+ \sum_{m \notin \{1, 2, ..., n\}} a_i a_k P_m$$

$$= \sum_{k=1}^{d} \left( a_i P_l - a_k P_k \right) + \sum_{m \notin \{1, 2, ..., n\}} a_k P_m.$$  

Thus, the form of the optimal solution is unique: for any $k \in \{1, 2, ..., d\}$, only one of the $q_{jk}$s is one, other $q_{jk}$s are all zeros.

Step 2. With privacy constraints:

As $\text{Var}[X_i]$ is monotonically increasing when $q_{lk} > \lambda_k$, and monotonically decreasing when $q_{lk} < \lambda_k$. The optimal solution (with privacy constraints) lies on the boundaries of the constraints: $e^{-} = \frac{\lambda_k}{q_{jk}}$, or $\frac{\lambda_k}{q_{jk}} = e^{-}$ (under 0 $< q_{jk}^i$).

$$\sum_{n=1}^{d} q_{jk} = 1; \forall j, k \in \{1, 2, ..., d\}.$$  

When one of the probabilities of $q_{mn1}, q_{mn2}, ..., q_{mnd}$ approaches 1 and others approaches 0, there are $d$ possible selections, and consider all the $m \in \{1, 2, ..., d\}$ there
are $d!$ feasible solutions. We now consider the case where $q_{kk}^i$ approach 1 for all $k \in \{1, 2, ..., d\}$ and other $q_{kk}^j$ are approaching 0. For the $q_{kk}^i$ which approach 1, the upper bounds is valid, and for $q_{kk}^j$ which approach 0, the lower bounds are valid. Considering the privacy constraints, we know the upper bound of $q_{kk}^i$ is $\lambda_k^i/e^\epsilon$ and the lower bound of $q_{kk}^i$ is $\lambda_k^i/e^\epsilon$.

We can check whether $q_{kk}^i$ are in the feasible region:

\begin{align}
\lambda_k^i - e^{-\epsilon} &= e^\epsilon P_k^i - 1 - e^{-\epsilon} \geq 0, \tag{45} \\
e^{-\epsilon} - \lambda_k^i &= e^{-\epsilon} - e^\epsilon P_k^i - 1 \geq 0. \tag{46}
\end{align}

So, when $q_{kk}^i$ reach the lower bound, $q_{kk}^i$ is still in the feasible region, it is readily to check that when $q_{kk}^i$ reaches the upper bound, $q_{kj}^i$ do not satisfy the privacy constraints.

**APPENDIX D**

**PROOF OF LEMMA 3**

*Proof.* As the MSE is the difference between the variance of the raw data and the variance of the estimator, when $d$ is fixed, the variance of the raw data is fixed, it is equivalent to show when $f \neq d$, the variance of the estimator decreases.

We know the optimal solution of the parameters of any input $X_i = a_k$ are in the form of: $q_{kk}^i$ is approaching 1 while other $q_{kj}^i$ are approaching 0 so that each input value can be inferred by a particular output. For example, given $Y_i = a_k$, one can probably infer that $X_i$ is also $a_k$ and the confidence increases with $\epsilon$.

- Assume that $d < f$, when $d$ is fixed, $\text{Var}(X)$ is also fixed. denote $\text{Var}(X_i)$ as the variance of the estimator with $d = f$ and $\text{Var}(X_i')$ as the variance of the estimator with $d > f$. Recall that

\begin{align}
\hat{X}_i &= \sum_{j=1}^{d} \sum_{k=1}^{d} a_j P_r(X_i = a_j|Y_i = a_k) I_k, \tag{47} \\
\hat{X}_i' &= \sum_{j=1}^{d} \sum_{k=1}^{d} a_j P_r(X_i = a_j|Y_i = a_k) I_k'. \tag{48}
\end{align}

First assume that for each $j \in \{1, 2, ..., d\}$, $k \in \{1, 2, ..., f\}$, the parameters of $\hat{X}_i$ and $\hat{X}_i'$ are identical. We know that for each $j \in \{1, 2, ..., d\}$, $k \in \{1, 2, ..., f\}$, $a_j P_r(X_i = a_j|Y_i = a_k) \geq 0$, thus $\text{Var}(X_i')$ is monotonically increasing with $f$.

Notice that the parameters of $\hat{X}_i$ and $\hat{X}_i'$ can not be identical as for at least one $j$, $q_{kj}^i$ will increase for $k \in \{f + 1, f + 2, ..., d\}$, $j \in \{1, 2, ..., f\}$. However, this will make each $P_r(X_i = a_k|Y_i = a_j)$ smaller, thus $P_r(X_i = a_k|Y_i = a_j) > P_r(X_i' = a_k|Y_i = a_j)$. As a result: $\text{Var}(X_i) > \text{Var}(X_i')$.

- Assume that $d < f$, this case can be viewed as a special case of the general model with $P_{k+1}^i = P_{\bar{k}+1}^i = ... = P_f^i = 0$. Thus the optimal solutions is straightforward: $q_{kk}^i = 1 - (1 - P_k^i)/e^\epsilon$, $q_{kj}^i = P_j^i/e^\epsilon$ for $k, j \in \{1, 2, ..., d\}; q_{kj}^i = 0$, for $k \in \{1, 2, ..., d\}, j \in \{d + 1, d + 2, ..., f\}$. As a result, the optimal solution is equivalent to the case of the general model with $d = f$.\[
\square
\]

**APPENDIX E**

**PROOF OF THEOREM 3**

*Proof.* The first step is to show the minimal MSE is achieved when $q_0$ and $q_1$ are at their minimum, which can be proved by taking derivative of the MSE function with respect to $q_s$ to show that MSE is increasing with $q_s$.

The second step is to find the minimum values of $q_s$, which are found according to the privacy constraints. To derive the monotonicity of the privacy metric with respect to $q_s$ Define $F_1 = \frac{P_r(g_i = g_i|Y_i = 1)}{P_r(g_i = g_i)}$, $F_2 = \frac{P_r(g_i = g_i|Y_i = 0)}{P_r(g_i = g_i)}$ which can be further expressed as

\begin{align}
F_1 &= \frac{\text{Pr}(Y_i = 0|g_i = g)}{\text{Pr}(Y_i = 0)} = \frac{(1 - q_0)T_0^g + q_1T_1^g}{(1 - q_1)T_0^g + (1 - q_0)(1 - P_1^g)}, \\
F_2 &= \frac{\text{Pr}(Y_i = 1|g_i = g)}{\text{Pr}(Y_i = 1)} = \frac{q_0T_0^g + (1 - q_1)T_1^g}{(1 - q_1)T_1^g + (1 - q_0)(1 - P_1^g)}. \tag{49}
\end{align}

Taking derivative over $q_0^i$ and $q_1^i$, we have:

\begin{align}
\frac{\partial F_1}{\partial q_0^i} &= \frac{(T_2^g - T_3^g)q_0^i - (q_1P_1^g + (1 - q_0)(1 - P_1^g))e^\epsilon}{(q_1P_1^g + (1 - q_0)(1 - P_1^g))e^\epsilon} = \frac{q_0^i}{(q_1P_1^g + (1 - q_0)(1 - P_1^g))e^\epsilon} = \frac{q_0^i}{(q_1P_1^g + (1 - q_0)(1 - P_1^g))e^\epsilon}, \\
\frac{\partial F_2}{\partial q_0^i} &= \frac{(T_3^g - T_2^g)q_0^i - (q_1P_1^g + (1 - q_0)(1 - P_1^g))e^\epsilon}{(q_1P_1^g + (1 - q_0)(1 - P_1^g))e^\epsilon} = \frac{q_0^i}{(q_1P_1^g + (1 - q_0)(1 - P_1^g))e^\epsilon}.
\end{align}

So we know, when $T_3^g > P_1^g$, $F_1$ is monotonically increasing with $q_s$, whereas $F_2$ is monotonically decreasing with $q_s$, so the minimum $q_s$’s are achieved when $F_1^i = e^{-\epsilon}$ and $F_2^i = e^\epsilon$. Solving the equations, and we get: $q_0^i = \frac{T_3^g - T_2^g}{(e^\epsilon + 1)(T_2^g - T_3^g)}; q_1^i = \frac{1 + T_3^g - e^{-\epsilon}}{(e^\epsilon + 1)(T_2^g - T_3^g)}$. When $T_3^g < P_1^g$, $F_1^i$ is monotonically decreasing with $q_s$, whereas $F_2$ is monotonically increasing with $q_s$, so the minimum $q_s$’s are achieved when $F_1^i = e^{-\epsilon}$ and $F_2^i = e^\epsilon$. Solving the equation, and we get:

\begin{align}
q_0^i &= \frac{P_1^g - T_2^g e^{-\epsilon}}{(e^\epsilon + 1)(P_1^g - T_2^g)}; q_1^i = \frac{1 + T_3^g - e^{-\epsilon}}{(e^\epsilon + 1)(T_3^g - P_1^g)}. \tag{50}
\end{align}

The final step is to test the value of $q_0^i$ and $q_1^i$ as functions of $T_3^g$, taking derivative on $q_s$’s, we have the first set of solutions are monotonically increasing with $T_3^g$, and the second set of solutions are monotonically decreasing with $T_3^g$. Thus to find a pair of $q_0^i$ and $q_1^i$ satisfy the $T_3^g$ for all $g \in G$, we take the maximum of all the possible values, and as $q_s$’s are non-negative, thus another candidate in the max function is 0.\[
\square
\]
APPENDIX F

PROOF OF COROLLARY 1

Proof. For the utility function of the problem of histogram estimation, the parameter-related term can be expressed as:

\[
N \sum_{i=1}^{N} \sum_{m=1}^{d} \left\{ \text{Var} \left( \sum_{k=1}^{m} \Pr(X_i = a_m | Y_i = a_k) \mathbb{1}(Y_i = a_k) \right) \right\}
\]

\[
= N \sum_{i=1}^{N} \sum_{m=1}^{d} \sum_{k=1}^{m} \Pr(X_i = a_m | Y_i = a_k)^2 \text{Var}(\mathbb{1}(Y_i = a_k))
\]

\[
+ N \sum_{i=1}^{N} \sum_{m=1}^{d} \sum_{k=1}^{m} \Pr(X_i = a_m | Y_i = a_k) \Pr(X_i = a_m | Y_i = a_l)
\]

\[
\cdot \text{Cov}(\mathbb{1}(Y_i = a_k), \mathbb{1}(Y_i = a_l))
\]

\[
= N \sum_{i=1}^{N} \sum_{m=1}^{d} \left\{ \sum_{k=1}^{m} \left( \frac{q_{mk} q_{ml}}{\lambda_k^2} - 1 \right) - \sum_{k=1}^{m} \sum_{l \neq k} q_{mk} q_{ml} (P_m^2) \right\}
\]

(50)

For the i-th user, taking partial derivative of Eq. (50) with respect to \(q_{mk}, \forall m \neq k \in \{1, 2, ..., d\} ; \)

\[
\frac{(\sum_{d \neq m} q_{mk})^2 (1 - P_m^2)(q_{mk} - \lambda_k^2)}{\lambda_k^2}
\]

(51)

where similar conclusion can be drawn as in the proof of Theorem 2: minimum is achieved when \(q_{mk} = \lambda_k^2 \); the function is monotonically decreasing when \(q_{mk} < \lambda_k^2 \) and increasing when \(q_{mk} \geq \lambda_k^2 \). Thus, the optimal solutions are found at the boundary of the privacy constraints. As the optimization problem defined in Eq. (28) is identical to that of Eq. (20), they also have the same optimal solutions. \(\square\)

REFERENCES

[1] T. Maddox, “The dark side of wearables: How they’re secretly jeopardizing your security and privacy,” 2016.

[2] J. Krumm, “Inference attacks on location tracks,” in Persuasive Computing (A. LaMarca, M. Langheinrich, and K. N. Truong, eds.), (Berlin, Heidelberg), pp. 127–143, Springer Berlin Heidelberg, 2007.

[3] K. Weaver, “How smart meters invade individual privacy,” 2014.

[4] P. Samarati and L. Sweeney, “Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression,” tech. rep., 1998.

[5] C. Dwork, “Differential privacy: A survey of results,” in Theory and Applications of Models of Computation: 5th International Conference, TAMC (M. Agrawal, D. Du, and Z. Duan, eds.), pp. 1–19, 2008.

[6] C. Dwork, “Differential privacy,” in Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Part II (M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, eds.), pp. 1–12, 2006.

[7] C. Dwork, F. McSherry, and K. Nissim, “Calibrating noise to sensitivity in private data analysis,” in Theory of Cryptography: Third Theory of Cryptography Conference, pp. 265–284, 2006.

[8] \text{FITZPATRICK}, “What to do after the massive yahoo hack,” 2016.

[9] R. Chen, H. Li, A. K. Qin, S. P. Kasiviswanathan, and H. Jin, “Private spatial data aggregation in the local setting,” in 2016 IEEE 32nd ICDE, pp. 289–300, May 2016.

[10] S. L. Warner, “Randomized response: A survey technique for eliminating evasive answer bias,” Journal of the American Statistical Association, vol. 60, no. 309, pp. 63–69, 1965.

[11] J. Freudiger, R. Shokri, and J-P. Hubaux, “Evaluating the privacy risk of location-based services,” in Proceedings of the 15th International Conference on Financial Cryptography and Data Security, FC’11, (Berlin, Heidelberg), pp. 31–46, Springer-Verlag, 2012.

[12] P. Kairouz, S. Oh, and P. Viswanath, “Extremal mechanisms for local differential privacy,” in Advances in Neural Information Processing Systems 27, pp. 2879–2887, Curran Associates, Inc., 2014.

[13] S. Xiong, A. D. Sarwate, and N. B. Mandayam, “Randomized requantization with local differential privacy,” in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2189–2193, March 2016.

[14] A. D. Sarwate and L. Sarkar, “A rate-disortion perspective on local differential privacy,” in 2014 52nd Annual Allerton Conference on Communication, Control, and Computing, pp. 903–908, Sept 2014.

[15] Ulfar Erlingsson, V. Pihur, and A. Korolova, “Rappor: Randomized aggregatable privacy-preserving ordinal response,” in Proceedings of the 21st ACM CCS, 2014.

[16] J. Tang, A. Korolova, X. Bai, X. Wang, and X. Wang, “Privacy loss in apple’s implementation of differential privacy on MacOS 10.12.” CoRR, vol. abs/1709.02753, 2017.

[17] T. Wang, J. Blocki, N. Li, and S. Jha, “Locally differentially private protocols for frequency estimation,” in 26th USENIX Security 17, pp. 729–745, USENIX Association, 2017.

[18] J. Blocki, N. Li, and S. Jha, “Context-aware data aggregation in the local setting,” in Electronic Notes in Theoretical Computer Science, vol. abs/1710.09549, 2017.

[19] R. Bassily, K. Nissim, U. Stemmer, and A. Thakurta, “Practical local private heavy hitters,” CoRR, vol. abs/1707.04982, 2017.

[20] C. Huang, P. Kairouz, X. Chen, L. Sankar, and R. Jagapath, “Context-aware generative adversarial privacy,” CoRR, vol. abs/1710.09549, 2017.

[21] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, “Context aware computing for the internet of things: A survey,” IEEE Communications Surveys Tutorials, vol. 16, pp. 414–454, First Quarter 2014.

[22] M. E. Andrés, N. E. Bordenabe, and K. Chatzikokolakis, “Geo-indistinguishability: Differential privacy for location-based systems,” CoRR, vol. abs/1212.1984, 2012.

[23] F. Tramèr and Z. Huang, “Differential privacy with bounded priors: Reconciling utility and privacy in genome-wide association studies,” in Proceedings of the 22Nd ACM SIGSAC, pp. 1286–1297, 2015.

[24] D. Kifer and A. Machanavajjhala, “No free lunch in data privacy,” in Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data, SIGMOD ’11, (New York, NY, USA), pp. 193–204, ACM, 2011.

[25] C. Dwork and G. N. Rothblum, “Concentrated differential privacy,” CoRR, vol. abs/1603.01887, 2016.

[26] C. Dwork and A. Roth, “The algorithmic foundations of differential privacy,” Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3-4, pp. 211–407, 2014.

[27] P. Cuff and L. Yu, “Differential privacy as a mutual information constraint,” CoRR, vol. abs/1608.03677, 2016.

[28] S. Asoodeh, F. Alajaji, and T. Linder, “Notes on information-theoretic privacy,” in 2014 52nd Allerton, pp. 1272–1278, Sept 2014.

[29] W. Wang, L. Ying, and J. Zhang, “On the relation between identifiability, differential privacy, and mutual-information privacy,” IEEE Transactions on Information Theory, vol. 62, pp. 5018–5029, Sept 2016.

[30] S. Asoodeh, F. Alajaji, and T. Linder, “On maximal correlation, mutual information and data privacy,” in IEEE CWSIT, pp. 27–31, July 2015.

[31] D. Kifer and A. Machanavajjhala, “A rigorous and customizable framework for privacy,” in Proceedings of the 31st ACM SIGMOD-SIGACT-SIGAI symposium on Principles of Database Systems, pp. 77–88, ACM, 2012.

[32] B. Yang, I. Sato, and H. Nakagawa, “Bayesian differential privacy on correlated data,” in Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, SIGMOD ’13, (New York, NY, USA), pp. 747–762, ACM, 2015.

[33] N. Li, W. Qardaji, D. Su, Y. Wu, and W. Yang, “Membership privacy: A unifying framework for privacy definitions,” in Proceedings of the 2013 ACM SIGSAC Conference on Computer & Network Security, CCS ’13, (New York, NY, USA), pp. 889–900, ACM, 2013.

[34] F. Calmon and N. Fawaz, “Privacy against statistical inference,” vol. abs/1608.03677, 2016.

[35] B. Jiang, M. Li, and R. Tandon, “Context-Aware data aggregation with localized information privacy,” in 2018 IEEE Conference on Communications and Network Security (CNS) (IEEE CNS 2018), (Beijing, P.R. China), May 2018.
[36] P. Kairouz, Oh, and Sewoong, “Extremal mechanisms for local differential privacy,” in Advances in Neural Information Processing Systems 27, pp. 2879–2887, 2014.

[37] Y. Wang, S. Song, and K. Chaudhuri, “Privacy-preserving analysis of correlated data,” CoRR, vol. abs/1603.03977, 2016.

[38] I. Issa, S. Kamath, and A. B. Wagner, “An operational measure of information leakage,” in 2016 Annual Conference on Information Science and Systems (CISS), pp. 234–239, March 2016.

[39] J. Lee and C. Clifton, “Differential identifiability,” in Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’12, (New York, NY, USA), pp. 1041–1049, ACM, 2012.

[40] B. Jiang, M. Li, and R. Tandon, “Local information privacy with bounded prior,” in ICC 2019 - 2019 IEEE International Conference on Communications (ICC), pp. 1–7, May 2019.

[41] F. A. S. Asoodeh and T. Linder, “Privacy-aware mmse estimation,” in 2016 IEEE ISIT, pp. 1989–1993, July 2016.

[42] Z. Qin, Y. Yang, and T. Yu, “Heavy hitter estimation over set-valued data with local differential privacy,” in Proceedings of the 2016 ACM SIGSAC, CCS ’16, pp. 192–203, 2016.

[43] T. Wang, N. Li, and S. Jha, “Locally differentially private frequent itemset mining,” in 2018 IEEE Symposium on Security and Privacy (SP), vol. 00, pp. 578–594.

[44] A. Papoulis and S. Pillai, Probability, random variables, and stochastic processes. McGraw-Hill, 2002.