Utility Cost of Formal Privacy for Releasing National Employer-Employee Statistics*

Samuel Haney[1], Ashwin Machanavajjhala[1], John M. Abowd[2], Matthew Graham[3], Mark Kutzbach[3], and Lars Vilhuber[2]

[1]Duke University,
[2]U.S. Census Bureau,
[3]Cornell University,

Abstract

Published tabular summaries of linked employer-employee data usually use a job frame (statutory employer linked to a specific employee) but include characteristics of both the individual (employee) and workplace (employer establishment). We show that state of the art differentially private algorithms add too much noise for the output to be useful. Instead, we identify the privacy requirements mandated by current interpretations of the relevant laws and formalize them with a model that simultaneously protects individuals and establishments using parameters that control the conventional differential privacy for individuals and a generalization that provides a similar privacy guarantee for the employment magnitudes associated with an employer establishment. We implement our model using three alternative noise distributions. We illustrate the system using the LEHD Origin-Destination Employment Statistics (LODES) database displayed in the U.S. Census Bureaus OnTheMap application; We present results for cross-sectional employment summaries for combinations of employer industry, geography, and ownership; and employee sex and educational attainment. Empirical evaluation of utility for these data shows that for reasonable values of the privacy-loss parameter epsilon at least 1, the additive error introduced by our provably private algorithms is comparable, and in some cases better, than the error introduced by existing statistical disclosure limitation (SDL) techniques that have no provable privacy guarantees. For some complex queries currently published, however, our algorithms do not have utility comparable to the existing traditional SDL algorithms. Those queries are fodder for future research.

1 Introduction

In this paper we present a case study in applying provably private algorithms for publishing tabular summaries of linked employer-employee data; i.e., data about business establishments and characteristics of their workforces. Such publications are used to compute national economic indicators including job creation and destruction statistics. Canada, the United Kingdom, and most Eurostat countries have regularly published data from establishment-based surveys that directly measure employee characteristics. Statistics Sweden explicitly publishes such tabulations based on its Business Database (a frame containing lists of workplaces)combined with establishment-based surveys that measure employee characteristics. In the U.S., the Census Bureau publishes County Business Patterns and Quarterly Workforce Indicators using establishment frames that include characteristics of employees.

This paper focuses on one such employer-employee dataset published by a national statistical agency. The dataset summarizes millions of jobs by a set of employer (workplace location, industry sector and ownership type), employee (age, gender, race, ethnicity and education) and job characteristics.

These linked employer-employee data cannot be released as unaltered establishment-level microdata. They are subject to legal confidentiality protections that apply to both the establishments and employees. In particular, statistical agencies are usually required by appropriate laws in their countries to protect whether or not a specific individual is employed at a workplace and the reported count of employees at that workplace. However, the existence of one or more employers in a location, industrial, or ownership type does not usually require privacy protection.

The combination of detailed geography (at the resolution of a very fine grid on a city's area) with characteristics of establishments and employees results in *sparse* tabulations for which many cells have only a few contributing establishments and/or individuals. In addition, the outcome being tabulated, employment, is highly *right skewed* (i.e., has many large outlying values) at the establishment level. The combined effect of this sparsity and skewness is the potential for re-identification attacks. Violating the statutory confidentiality pledge, for example by publishing data that permit re-identifying employees or inferring exact employer characteristics, can result in fines and potential imprisonment for employees of statistical agencies who authorize these publications. Thus, privacy must be ensured as a matter of criminal law. Even more important to most statistical agencies is the loss of reputation and concomitant degradation of response rates and data quality that would ensue from breaches of the confidentiality protection laws. With the stakes so high, we ask:

- "Can we develop algorithms for releasing such data that provably uphold the privacy requirements mandated by law?", and
- "What is the loss of data usefulness (utility cost) when re-

*Disclaimer: Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Machanavajjhala acknowledges support from NSF grants 1253327, 1408982, and 1443014. Abowd acknowledges direct support from NSF grants BCS-0941226 and TC-1012593. Abowd and Vilhuber acknowledge support from NSF grant SES-1131848.

leasing summaries that provably ensure privacy?"

Statutory requirements clearly require data publication in summary format while prohibiting statistical agencies from publishing the raw microdata. The laws do not, however, prescribe exactly how to publish the data [12]. Current publications of employment counts for detailed industries and geographies use a variety of confidentiality protection methods known collectively as *statistical disclosure limitation* (SDL) to limit potential re-identification attacks. These SDL methods are the *de facto* interpretations of the legally required confidentiality protections [11, 18]. For instance, in the data we consider, workplace counts are protected by an input noise infusion system that provably avoids exact disclosures of establishment-level employment counts but has no guarantee that informed attackers cannot infer actual establishment job counts to within some stated level of accuracy [1, 2, 16]. We present examples of such attacks in Section 5.2.

To overcome the aforementioned limitations of the current state of the art, we consider *ϵ-differential privacy* [14, 13], which is considered the leading standard for releasing data with formal privacy guarantees. A data release algorithm satisfies differential privacy if, for all allowable input databases, its output is insensitive to the presence or absence of any one sensitive record in the input database. This translates into privacy semantics [23] that an informed attacker's odds of whether a specific record was in or out of the database change at most by a factor of e^{ϵ} after seeing the output of a differentially private algorithm. ϵ is called the privacy-loss parameter, with smaller values resulting in more privacy. However, there is no known work linking the privacy guaranteed by differential privacy to any explicit statutory confidentiality requirements. Additionally, applying standard variants of differential privacy to protect establishment job counts results in output noise infusion that completely destroys all the signal in the data (due to the extreme right skewness in the data). Hence, we make the following contributions:

- We review the legal regulations that national statistical agencies are mandated to follow when releasing employer-employee data and identify privacy requirements implied by their current interpretation (Section 4). We mathematically formalize these requirements using the Pufferfish privacy framework [23, 24]. These requirements ensure that a strongly informed attacker cannot (a) infer whether an employee held a job, and (b) learn establishment sizes to within a pre-specified multiplicative factor from the output of a data release.
- While SDL techniques in use avoid exact disclosures, we show that these techniques do not satisfy our privacy requirements using examples of unbounded inferential disclosures that are possible using current releases (Section 5). We also show that standard variants of differential privacy either do not satisfy the privacy requirements or result in an extreme loss in utility (Section 6).
- We formulate novel privacy definitions that satisfy the formal privacy requirements, and provably limit an informed attacker's ability to make inferences about employees and employers (Section 7). Thus, our algorithms are able to uphold current interpretations of the laws governing these data.
- We develop algorithms for releasing counts that satisfy our new privacy definitions. We prove analytical bounds on the errors for each algorithm (Sections 8 and 9).
- We empirically evaluate the data quality cost of provable privacy using the production data from an employer-employee dataset maintained by a national statistical agency (Section 10). For releasing tabular summaries and rankings, we compare the error introduced by our algorithms to the error produced by the existing protection scheme. We show that we can release tabular summaries of establishment characteristics with additive error that is comparable (within a factor of 3) and in some cases smaller than the error introduced by the current SDL techniques (with no provable privacy guarantees) for reasonable values of the privacy-loss parameters. For counts and rankings, the relative error of our new algorithms is comparable to the error induced by the existing SDL methods. Since tabular summaries of employment by establishment characteristics, such as the Business Database statistics released in Sweden or the County Business Patterns released in the U.S., are important publications, it is significant that we are able to achieve provable privacy protection with relatively little sacrifice in data utility as compared to existing methods.
- For tabulations involving both establishment and employee characteristics, under a weaker (yet formal) privacy guarantee, we are able to show competitive error for releasing single queries and rankings. Our algorithms experience larger errors (within a factor of 100 compared to existing methods) for releasing tabular summaries involving multiple employee characteristics. Understanding whether the magnitude of this error is fundamental to provably private algorithms or whether better algorithm design could lower such errors is an avenue for future work.

We next introduce differential privacy in Sections 2 and the data we work with in Section 3.

2 Preliminaries

Database and Queries Let D be a table of records with schema $(A_1, \ldots, A_k)$. The domain of each attribute A_i is denoted $dom(A_i)$. For a set of attributes $V = \{A_{i_1}, \ldots, A_{i_m}\}$, let $dom(V)$ represent the multidimensional domain $\times_{A \in V} dom(A)$. For each record t in the table, we let $t[A_i] \in dom(A_i)$ be value of attribute A_i. Let $n = |D|$ denote the size of the table; i.e., D has n records. A database with schema $(S_1, \ldots, S_m)$ is a collection of tables $(D_1, \ldots, D_m)$, where D_i has schema S_i.

We will consider *marginal queries* over tables in this paper.

Definition 2.1 (Marginal Query)**.** *Let $V = \{A_{i_1}, \ldots, A_{i_m}\}$ denote a subset of attributes chosen from D. Let $dom(V) = \times_{A \in V} dom(A)$. The marginal query $q_V(D)$ is defined as a vector of $|dom(V)|$ counts, one for each* cell $\mathbf{v} = (v_1, \ldots, v_m) \in dom(V)$. *The count corresponding to cell $\mathbf{v}$, denoted by $q_V(D, \mathbf{v})$ is*

$$|\{t \in D \mid t[A_{i_1}] = v_1 \wedge \ldots \wedge t[A_{i_m}] = v_m\}| \qquad (1)$$

$q_{\emptyset}(D)$ returns a single cell whose count is the size of the table.

The marginal query can be succinctly expressed in SQL as:

Select Count(*) From D Group By $A_{i_1}, \ldots, A_{i_m}$

Differential Privacy A mechanism or algorithm is differentially private if its output is not significantly affected by presence or absence of a single record from the input table. Let D and D' be tables that differ in the presence of a single record; i.e., $|(D \setminus D') \cup (D' \setminus D)| = 1$. We call such tables *neighbors*.

Definition 2.2 ((ϵ, δ)-Differential Privacy [15])**.** *Let $\mathcal{M}$ be a randomized algorithm. Let the tables D and D' be neighbors with the same schema. Then $\mathcal{M}$ satisfies (ϵ, δ)-differential privacy if for all D and D' and for all $S \subseteq range(\mathcal{M})$,*

$$\Pr[\mathcal{M}(D) \in S] \;\leq\; e^{\epsilon} \cdot \Pr[\mathcal{M}(D') \in S] + \delta$$

δ allows for the ratio of probabilities to be unbounded with a small failure rate. Values of δ that are $\Omega(1/n)$ should be avoided, since algorithms that release more than a constant number of unperturbed records from the database satisfy such a definition. When $\delta = 0$, we refer to the condition as ϵ-differential privacy.

Queries over tables can be answered while satisfying differential privacy by adding noise that is related to the *sensitivity* of the query.

Definition 2.3 (Sensitivity)**.** *Let $\mathcal{I}$ denote the set of all tables with a given schema. Let $q : \mathcal{I} \rightarrow \mathbb{R}^d$ be a query function on that table that outputs a vector of d real numbers. The sensitivity of q, denoted Δ_q, is*

$$\Delta_q = \max_{D, D' \text{ neighbors}} ||q(D) - q(D')||_1.$$

The Laplace mechanism is a commonly used ϵ-differentially private technique.

Definition 2.4 (Laplace Mechanism [15])**.** *Let $q : \mathcal{I} \rightarrow \mathbb{R}^d$ be a query on a table. Let $\eta \sim Lap(\lambda)$ denote a random variable drawn from the Laplace distribution with pdf $Pr[\eta = x] \propto e^{-|x|/\lambda}$. The algorithm which returns $\tilde{q}(D) = q(D) + \eta^d$ satisfies ϵ-differential privacy, where η^d is a vector of d independently drawn Laplace random variables.*

Definition 2.5 (Expected L_p Error)**.** *Let $q : \mathcal{I} \rightarrow \mathbb{R}^d$ be a query over a table, and $\tilde{q}(D)$ be the noisy answer returned by an algorithm. The expected L_p error of the algorithm is:*

$$\mathbb{E}\left(||q(D) - \tilde{q}(D)||_p\right) \tag{2}$$

where $||x||_p$ is the L_p norm, and expectation is over the randomness of the algorithm.

We will use the expected L_1 error to quantify the utility of algorithms.

The privacy loss increases when multiple queries are answered on the database, and we reason about this loss for differential privacy using the following composition rule:

Theorem 2.1 (Sequential Composition)**.** *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be ϵ_1- and ϵ_2-differentially private algorithms. Releasing the outputs of $\mathcal{M}_1(D)$ and $\mathcal{M}_2(D)$ on the same input D results in $(\epsilon_1 + \epsilon_2)$-differential privacy.*

Thus, the privacy-loss parameter is often called the *privacy budget* – the analyst is allowed to pose multiple queries as long as the total privacy loss from answering all queries is no greater than ϵ. In other words, a privacy-loss equal to ϵ exhausts the privacy budget.

3 The Dataset

The employer-employee data we use in this paper are produced by a national statistical agency by composing administrative records, census and survey data focused on the labor market, worker, and establishment statistics. These data are published as summaries of an annual cross-section of jobs held on a specific date of each year. We consider one such cross-section.

3.1 Table and Database Structure

The data are organized as a relation with three database tables – Workplace, Worker and Job. The Workplace table contains one record per establishment and describes the following attributes – industry (denoting the industry in which the establishment operates), ownership (public/private), and geography (the place where the establishment is located). The Worker table contains one record for each individual working in any establishment at that point of time. Worker attributes include age, sex, race, ethnicity, and education. Finally, the Job table contains pairs (w, i) of worker and workplace IDs denoting that worker i works at establishment w. We assume each worker has exactly one job (although the original dataset does include secondary jobs). We have not documented some of the attributes that do not feature in our queries.

We support *marginal queries* that output counts of employment of the current year's cross section of jobs over workplaces where those establishments have been stratified by subsets of the available characteristics of employers and workers. Let V_I denote a subset of the worker attributes, and V_W denote a subset of workplace attributes. Let D denote the universal relation constructed by joining the Job table with the Worker and Workplace tables using the worker and workplace IDs, respectively. We call this the WorkerFull table as it contains one record for every record in Worker (each worker has exactly one job). The records in WorkerFull contain all the worker and workplace attributes. Records that share the same workplace ID represent the workforce of that establishment.

Every cell $(\mathbf{v}_I, \mathbf{v}_W)$ in the marginal query $q_{V_I \cup V_W}(D)$ (as in Definition 2.1) represents the number of workers matching the criteria in $\mathbf{v}_I$ who work in establishments matching the criteria in $\mathbf{v}_W$. The output is employment counts stratified by employee attributes in V_I and workplace attributes V_W.

3.2 Usage Scenarios

Researchers and analysts working with these data request answers to one or more marginal queries and use these summary tabulations as inputs to further data analysis. In this paper, we focus on analyzing the privacy and utility tradeoffs from releasing the output to one marginal query. Analyzing the privacy of multiple queries is a straightforward application of Theorem 2.1 since our privacy definitions also satisfy sequential composition like differential privacy. Note that the examples of queries and protection parameters provided here are meant to illustrate our methods. We have made no recommendation to the national statistical agency regarding the set of queries, methods, and privacy parameters to use for a full implementation.

We highlight two application scenarios that use summary tables computed from employer-employee datasets.

Resource Allocation: Employer-employee data are used for a number of planning and development purposes. These counts are used to decide whether or not a specific amount of funding can make a sufficient per capita impact in a given geographical area (say a county or a city), and decide which areas receive the funding. Errors in counts would affect whether or not an area received funding thus motivating our use of L_1 error as a measure of data accuracy. Experiments in Section 10 explore the effect of privacy parameters on L_1 error.

Ranking: Users are often interested in comparing counts across a list of units. As a well-known example, magazines and news outlets regularly publish rank-ordered national lists of cities by various attributes within city-size classes. Users currently interact with our employer-employee data through online tools that encourage them to rank job counts of a series of areas from largest to smallest. For example, a business might be interested in the ranked order of places (e.g., cities, or towns) by job count, within a state, for deciding where to open a new establishment. Exercises in Section 10 use Spearman's rank-order correlation to explore the accuracy of ranked lists produced by our algorithms as compared to the ordering produced by the current SDL scheme.

4 Privacy Requirements

We derive our privacy requirements from relevant national laws pertaining to confidential data collected for statistical purposes from employers or employees. We first discuss the general privacy requirements, and then mathematically formalize them using the Pufferfish framework.

4.1 Generic Legal Environment

Information on workers and firms is protected by laws mandating confidentiality protections for individual and business information collected by the national statistical agency.[1] Under these laws, the agency may not make any publication that identifies the data provided by any particular establishment or individual. The responsible committee at the national statistical agency approves the release of data that, in its view, satisfy these statutory confidentiality-protection requirements.

The employer-employee data are derived from both employer (business-level) and job (individual-level) data, and thus, both legal protections apply. Over the years, the agency has interpreted the statute to require the following set of protections.

- The existence of a job held by a particular individual is confidential and must not be disclosed.
- The existence of an employer business as well as its type (or industry) and location is not confidential.
- The data on the operations of a particular business must be protected. In our context, that means that characteristics of an establishment's workforce (e.g., total employment and all disaggregations like number of female employees of age 20-25) must be protected.

[1] Precise laws are not cited due to double blind submission requirements. Many national statistical agencies have privacy requirements consistent with the United Nations' Fundamental Principles of Official Statistics, Principle 6.

Appendix A discusses how these statutory confidentiality requirements have been interpreted over the past half a century, and the development of SDL techniques that uphold them. However, none of the prior interpretations or techniques have attempted proving formal statements about the privacy that is guaranteed to individuals or businesses. In fact, we show (Section 5) that the SDL techniques currently used to publish employer-employee data may allow an informed attacker to infer confidential properties like the exact total employment of an establishment, or whether a certain employee is employed by a specific employer. We next present our formalization of the privacy requirements.

4.2 Formal Privacy Desiderata

We propose to design algorithms for releasing counts from employer-employee data that can provide provable guarantees of the following privacy desiderata. Our discussion focuses on formulating privacy requirements for a *one-time release* by the statistical agency using an algorithm denoted by A. Our privacy notions handle multiple releases through composition rules. We model the requirements based on the Pufferfish privacy framework [23, 24].

Informed Attacker: National statistical agencies are concerned about two kinds of attackers – uninformed and informed. Uninformed attackers can access the output of the algorithm A, but may not possess detailed background knowledge about specific individuals and establishments in the data. Informed attackers are more powerful. They possess specific knowledge about individual employees or employers, or statistics about those in the dataset. Examples of such attackers include a group of employees who would like to determine a private attribute of their co-worker, or one (or more) employer(s) attempting to learn detailed statistics about a competing employer. Our goal is to ensure the confidentiality of employer and employee characteristics from such attackers.

We assume the adversary knows the set of all establishments (say $\mathcal{E}$), and their public attributes (location, industry code and ownership). The attacker also knows the universe of all workers U. Each worker $w \in U$ has a set of private attributes $A_1 \ldots A_k$ (like age and sex). We add another attribute with domain $\mathcal{E} \cup \perp$ that represents whether w works in one of the establishments in $\mathcal{E}$, or not.

For each employee w, the attacker's belief is defined as π_w, a probability distribution over all the values in $\mathcal{T} = (\mathcal{E} \cup \perp) \times A_1 \times A_2 \times \ldots \times A_k$. $\theta = \prod_{w \in U} \pi_w$ represents the adversary's belief about all employees in the universe U. That is, the adversary possesses no knowledge correlating employees. We denote by $\Theta = \{\theta\}$, the set of all possible adversarial beliefs that assume no correlations between employees and between employers. Nevertheless, Θ includes informed attackers who may know exact information about all but one employee, and those who know exact information about all but one employer. We note that Θ contains very strong attackers. Algorithms that can provably protect against such attackers while ensuring error comparable to current SDL techniques would underscore the possibility that provable privacy could be achieved at low utility cost.

We distinguish a subset of attackers $\Theta_{weak} \subset \Theta$ as *weak* attackers. Weak attackers have no prior knowledge over worker

attributes – i.e., all workers are the same in their eyes. The weak attacker may still have the same detailed knowledge about establishments as our general attacker. We capture a weak adversary by requiring that the prior for each worker π_w be a product of $\pi_{(1,e)}$ (worker independent prior over establishments), and $\pi_{(2,w)}$ (a uniform prior over all worker attributes). We use these definitions to define a weaker privacy notion.

What should we protect: We now define what properties of the data we need to protect against such adversaries.

1. No re-identification of individuals: We would like to ensure that adversaries do not learn too much additional information about any single employee in the dataset when an algorithm A operates on the dataset D. In particular, they should not be able to determine (i) whether or not an employee is in or out of the dataset ($\perp$ versus not), (ii) whether or not an employee works at a specific (type of) employer (E versus $\mathcal{E} - E$, where E is a set of employers), and (iii) whether or not the employee has certain characteristics (e.g., Hispanic with age greater than 35).

We formalize this as follows. For any pair of values $a, b \in \mathcal{T}$, we want the ratio of the adversary's posterior odds (after seeing the output $A(D)$) that a worker record takes the value $w = a$ vs $w = b$ not to be too different from the adversary's prior odds that $w = a$ vs $w = b$. That is, we want to bound the Bayes factor: the ratio of the posterior odds to the prior odds.

Definition 4.1 (Employee Privacy Requirement)**.** *For randomized algorithm A, if for some $\epsilon \in (0, \infty)$, and for every employee $w \in U$, for every adversary $\theta \in \Theta$, for every $a, b \in \mathcal{T}$ such that $Pr_\theta[w = a] > 0$ and $Pr_\theta[w = b] > 0$, and for every output $\omega \in range(A)$:*

$$\log\left(\frac{Pr_{\theta,A}[w = a|A(D) = \omega]}{Pr_{\theta,A}[w = b|A(D) = \omega]} \Big/ \frac{Pr_\theta[w = a]}{Pr_\theta[w = b]}\right) \leq \epsilon \quad (3)$$

Then the algorithm A protects employees against informed attackers at privacy level ϵ.

Definition 4.1 bounds the logarithm of the maximum Bayes factor an informed attacker can achieve. This implies, as a consequence of the general bound on privacy-loss, that an informed attacker can't learn any property of a worker record with probability 1 after seeing the output of the algorithms unless the attacker already knew that fact, as reflected in his prior odds.

2. No precise inference of establishment size: An *informed* attacker should not *infer* the total employment of a single establishment to within a multiplicative factor of α. We do not require stronger privacy of the form "presence of an establishment must not be inferred," since (a) the existence of an employer establishment is considered public knowledge, (b) the data are an enumeration of all employer establishments, and (c) whether or not an establishment is big or small is well known. This requirement balances the legal need for protecting the operations of a business with widespread knowledge of approximate employment sizes of establishments.

We can formalize the employer-size privacy requirement as follows. For any establishment e, let $|e|$ denote the random variable representing the number of workers employed at e. We define the requirement for both informed and weak adversaries.

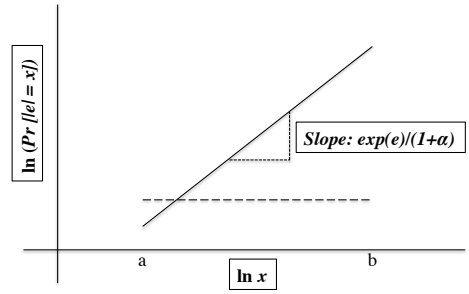

Figure 1: Pictorial representation of an attacker's prior and posterior belief about an establishment's size ($|e|$) permitted by Definition 4.2. The prior is $uniform(a, b)$ (dotted line). Note the log scale on both x and y axes.

Definition 4.2 (Employer Size Requirement)**.** *A randomized algorithm A protects establishment size against an informed attacker at privacy level (ϵ, α) if, for every informed attacker $\theta \in \Theta$, for every pair of numbers x, y, and for every output of the algorithm $\omega \in range(A)$,*

$$\left|\log\left(\frac{Pr_{\theta,A}[|e| = x|A(D) = \omega]}{Pr_{\theta,A}[|e| = y|A(D) = \omega]} \Big/ \frac{Pr_\theta[|e| = x]}{Pr_\theta[|e| = y]}\right)\right| \leq \epsilon \quad (4)$$

whenever $x \leq y \leq \lceil(1 + \alpha)x\rceil$ and $Pr_\theta[w = x], Pr_\theta[w = y] > 0$. We say that an algorithm weakly protects establishments against an informed attacker if the condition above holds for all $\theta \in \Theta_{weak}$.

As in Definition 4.1, this definition bounds the maximum Bayes factor the informed attacker can learn within the universe of allowable data tables. Unlike the case of individuals, Definition 4.2 does allow an adversary to learn about the gross size of an employer establishment. Consider the example illustrated in Figure 1. Suppose that the informed attacker has a uniform prior for the size of an establishment within the range $[a, b]$. This prior is represented as a dotted horizontal line in the figure. Definition 4.2 requires that the adversary's posterior odds be bounded for pairs x and y within a multiplicative factor of $(1 + \alpha)$. That means, the attacker's posterior could be a peaked distribution (shown as a solid line in the figure). The only constraint on the posterior beliefs is that a multiplicative increase in establishment size of $(1 + \alpha)$ should result in a multiplicative increase or decrease in the attacker's posterior belief of at most e^ϵ (as shown by a slope of $\epsilon/\ln(1 + \alpha)$ in Figure 1, where the axes are on a logarithmic scale).

3. No precise inference of establishment shape: An *informed* attacker cannot precisely infer the composition of a single establishment's workforce (e.g., the fraction of males who have a bachelor's degree or the fraction with Hispanic ethnicity). We call the distribution of an establishment's workforce based on worker characteristics its *shape*. One can think of this requirement as protecting the distribution of characteristics of the workforce, whereas the previous requirement protected the magnitude of each characteristic. We believe this shape requirement implements the legally mandated confidentiality of an establishment's operating characteristics. In Section 5.2, we show that an algorithm that releases the precise shape of the workforce allows an informed attacker to learn the establishment size as well. The

Name	Does this satisfy requirement on		
	Individuals	Employer Size	Employer Shape
Input Noise Infusion (Sec. 5)	No	No	No
Differential Privacy (individuals, Sec. 6)	Yes	No	No
Differential Privacy (establishments, Sec. 6)	Yes	Yes	Yes
Employer-Employee-privacy (Sec. 7)	Yes	Yes	Yes
Weak Employer-Employee-privacy (Sec. 7)	Yes	Yes*	Yes

Table 1: Privacy definitions and the requirements they satisfy. * means the the privacy requirements is satisfied under weak adversaries.

definition bounds the maximum Bayes factor that the informed adversary can learn within the set of allowable inputs.

Definition 4.3 (Employer Shape Requirement)**.** *Let $e_{\mathcal{X}}$ denote the subset of employees working at e who have values in $\mathcal{X} \subset A_1 \times \ldots \times A_k$. A randomized algorithm A protects establishment shape against an informed attacker at a privacy level of (ϵ, α), if for every informed attacker $\theta \in \Theta$, for every property of a worker record $\mathcal{X} \subset A_1 \times \ldots \times A_k$, for every pair of numbers $0 < p \leq q \leq \min(1, (1+\alpha)p)$, for every output of the algorithm $\omega \in range(A)$, and for every number z,*

$$\log \frac{Pr_{\theta,A}[|e_{\mathcal{X}}|/|e| = p, |e| = z | A(D) = \omega]}{Pr_{\theta,A}[|e_{\mathcal{X}}|/|e| = q, |e| = z | A(D) = \omega]} \frac{Pr_{\theta}[|e_{\mathcal{X}}|/|e| = p, |e| = z]}{Pr_{\theta}[|e_{\mathcal{X}}|/|e| = q, |e| = z]} \leq \epsilon \quad (5)$$

whenever $Pr_\theta[|e_{\mathcal{X}}|/|e| = p, |e| = z], Pr_\theta[|e_{\mathcal{X}}|/|e| = q, |e| = z] > 0$.

Table 1 summarizes whether or not certain SDL and formal privacy methods satisfy our privacy desiderata. The following sections provide definitions and analysis required to interpret this table.

5 Current SDL Protection

We first discuss how employer-employee data are protected using SDL techniques, and how they avoid exact disclosures in Section 5.1. However, these techniques leave the door open for inference attacks, a few of which are outlined in Section 5.2.

5.1 Input Noise Infusion

A popular technique for protecting employer-employee data is *input noise infusion* [1, 3], where the database is perturbed before answering queries. Every establishment w is assigned a unique distortion factor f_w, bounded away from 1. The unique, time-invariant, confidential distortion factor f_w, is within the union of the ranges $[1-\beta, 1-\alpha] \cup [1+\alpha, 1+\beta]$. The parameters $0 < \alpha < \beta$ are kept confidential in order to limit inference attacks. Zero counts are left unmodified.

More formally, consider a table WorkplaceFull that has one row per workplace w, as well as a histogram $\mathbf{h}(w)$ of counts of workers employed at w cross-tabulated over all combinations of worker attributes. Let c denote one of the cells (combinations of worker attributes like males, age 16-18, Hispanic, etc.), and let $\mathbf{h}(w, c)$ denote the count for workplace w in cell c. Counts in $\mathbf{h}(w)$ are perturbed to get $\mathbf{h}^\star(w)$ as follows.

$$\mathbf{h}^\star(w, c) \;=\; f_w \cdot \mathbf{h}(w, c) \quad (6)$$

To limit re-identification of individual workers, additional output perturbation is employed for small counts. Specifically, when a marginal query q_V is posed to the system (say employments counts tabulated by age, sex, and place), both the true answer $q_V(D)$ as well as a noise infused answer $q_V^\star(D)$ are computed. The latter is constructed by adding up appropriate counts from $\mathbf{h}^\star(w)$ for the establishments that satisfy the workplace criteria. If for a cell $\mathbf{v}$ in the output, the true count $q_V(D, \mathbf{v})$ lies within $(0, S)$, then the noise infused answer $q_V^\star(D, \mathbf{v})$ is replaced by a sample drawn from a posterior predictive distribution that always outputs integers $1, ..., \lfloor S \rfloor$. The small cell limit S is set to 2.5 for our dataset. Note that zero counts are unperturbed.

Privacy Properties

- *No Exact Disclosures about Establishments:* As a direct consequence of the gap around 1 in the distortion factor $f_w \in [1-\beta, 1-\alpha] \cup [1+\alpha, 1+\beta]$, an establishment's actual employment count is never used in any computations that produce tabular summaries. Hence, even if some cell count in a marginal query contains only one establishment, its employment count is not exactly revealed. The statutory requirement not to publish exact data about establishments is fulfilled by using the distortion factors α and β.
- *No Re-identification of Employees without Background knowledge:* Given the output of a single marginal query, an adversary can not re-identify the presence of a specific worker without additional background knowledge. This is ensured by replacing small counts using draws from a different distribution.

Nevertheless, the aforementioned scheme is vulnerable to inference attacks, especially in the presence of background knowledge, as discussed next.

5.2 SDL Vulnerabilities

The following two properties of the input noise infusion scheme allow inference about individuals, establishment sizes and establishment shapes:

- The same distortion factor f_w is used to perturb all the cells counts $\mathbf{h}(w)$ for an establishment w.
- If $\mathbf{h}(w) = 0$, then $\mathbf{h}^\star(w) = 0$.

We first note that the privacy requirement on establishment shape (Definition 4.3) is not satisfied. Consider a marginal query $q_{V_I \cup V_W}$, where V_I are attributes of the employee and V_W are attributes of the establishment. Suppose there is one combination $\mathbf{v}_W \in dom(V_W)$ such that exactly one workplace w fits that criterion.[2] Thus, counts output by the marginal query for

[2]The number of establishments in a cell are not published for the dataset we consider in this paper. However, there are combinations of V_W that contain only one workplace, and an adversary could know this.

all cells $(\mathbf{v}_W, c)$, for all $c \in dom(V_I)$ would represent employment counts for a single workplace. Whenever all these cell counts are greater than the small cell limit S, they are precisely the true count multiplied by the same (but unknown) noise factor f_w. This reveals the shape exactly.

Next, let us consider the privacy requirement on establishment size (Definition 4.2). Again consider the case when there is one combination $\mathbf{v}_W \in dom(V_W)$ such that exactly one workplace w fits that criterion. Additionally, suppose the attacker knows one of the cell counts $(\mathbf{v}_W, c)$ truthfully (say, the attacker knows there are a 100 males, age 20-25). If this count is greater than the small cell limit S, then adversary can reconstruct the noise factor f_w. This coupled with the fact the shape is revealed exactly, the attacker can reconstruct the counts in all other cells as well as the total size of the establishment's workforce.

The existence of this attack is understood by the agencies that use input noise distortion as a method of SDL. The attack is not currently considered a violation of the relevant data protection statutes because exact disclosure occurs only when one of the counts $\mathbf{h}(w, c)$ are known exactly for some cell c. The agencies assume that the only users who possess exact information on c are employees of a business that reported the data. If those employees are obligated to keep such information confidential, then it is the employer's duty to prevent the attack or prosecute the attacker. If they are not, then the data item itself is no longer confidential, and statutory protection doesn't apply because the employer released the value.

Next, we show that individual employees could be re-identified by informed attackers (thus violating privacy requirement in Definition 4.1). Suppose again a marginal query $q_{V_I \cup V_W}$ with one combination $\mathbf{v}_W \in dom(V_W)$ that fits exactly one workplace w. Additionally, suppose an adversary knows that there is only one employee in w with a college degree. If V_W contains the education attribute, then the only cells that correspond to having a college degree in $\mathbf{h}(w, c)$ with positive counts are those that correspond to the true values of the other attributes for that employee. Since zero counts are preserved in the algorithm described in the previous section, the attacker can infer the other true attributes for this employee looking at the published counts. Current publications of the employer-employee data we consider are vulnerable to this attack. Such attacks will be thwarted by the algorithms we propose.

6 Applying Differential Privacy

In this section we consider directly applying differential privacy to our problem. One can think of the entities in our problem (employers and employees) as nodes in a *bipartite* graph connected by edges that represent jobs. There is extensive work on applying differential privacy to graphs, which can be brought to bear on our problem [8, 10, 21, 20].

Two standard variants of differential privacy considered in the context of graphs are *edge* and *node* differential privacy. Edge differential privacy considers neighboring graphs that differ in the presence of a single edge. In our context that corresponds to adding or removing a single worker (or a job) from our database. We can show that this definition is sufficient to satisfy the employee privacy requirement (Definition 4.1). However, edge differential privacy does not ensure privacy of establishments (Definitions 4.2 and 4.3) (see appendix for proofs). For instance, under this definition an adversary is allowed to compute the number of workers at a single establishment satisfying any property by perturbing the true count with additive noise drawn from $Lap(1/\epsilon)$. We can show that the noise added is at most $\frac{\log(1/p)}{\epsilon}$ with probability $1 - p$ (i.e., at most 5 for $\epsilon = 1$ and $p = 0.01$). Knowing that the total employment in an establishment is 10,000 ± 5 is as good as knowing the true count.

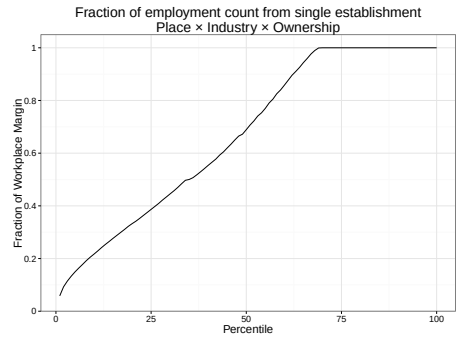

Figure 2: Employment counts from a single establishment

Node differential privacy considers neighboring graphs that differ in the presence of an single node and all the edges incident to it. In our context that corresponds to removing or *adding* a single employer along with all the workers who are employed at this employer. Node differential privacy is much stronger, and in the context of our problem will satisfy the employee and employer privacy requirements (Definitions 4.1, 4.2 and 4.3). However, this comes at a huge cost to utility.

There are two classes of techniques for answering aggregate queries on graphs – adding noise to true counts, and projection. The first class of techniques [20] adds noise to an aggregate count (e.g., number of employees in New York) with scale that is at least the size of the largest establishment that contributes to that count (this is sometimes called the *down sensitivity*, and is a weak lower bound on the noise scale required to ensure privacy). Thus the minimum relative error in such a count is at least the fraction of the aggregate contributed by a single establishment. For instance, we computed the tabular summary of counts by place, industry and ownership, where each cell represents the total employment in a specific place, for establishments with a specific industry code and ownership type. We also computed for each cell the fraction of the count contributed by a single establishment (or the ratio between the down sensitivity and the cell count). Figure 2 shows the CDF of these fractions. We can see that at least 50% of the cells have a single establishment contributing to more than 65% of the count, and would incur at least that much relative error if perturbed using noise scaled even by down sensitivity.

The projection techniques [8, 10, 21] modify the graph by adding or deleting edges and nodes until the maximum degree of a node is bounded by a small number θ. Queries on this bounded degree graph have bounded sensitivity (of θ) and hence can be answered by adding noise from $Laplace(\theta/\epsilon)$. However, such a procedure would severely distort the characteristics of employers with large sizes. Preserving properties of these

establishments is important for economic studies. For instance, the truncation method [21] projects the graph by removing nodes until all nodes have degree less than θ. Using this technique on our data with $\theta = 1000$ in the context of employer-employee data results in removing all establishments with size at least a 1000; between 740 and 815 establishments would be removed.[3] Moreover, in a tabular summary of counts by place, industry and ownership, over 93% of the counts have a total count less than a 1000. Adding Laplace noise with sensitivity of a 1000 (even with $\epsilon = 1$) would result in expected noise greater than the cell counts. Thus, projection techniques would destroy utility.

7 Formal Privacy Definition

We present two privacy definitions that ensure protection against informed and weak adversaries, respectively, as we have defined them. Our new definitions are in Section 7.1; their privacy semantics are in Section 7.2. Our formal privacy definitions ensure that the requirements in Section 4 are satisfied.

7.1 Privacy for Employee-Employer Data

We denote the set of establishments as $\mathcal{E}$ and the universe of workers as U.

Definition 7.1 (Strong α-Neighbors)**.** *Let D and D' be two employer-employee tables such that they differ in the employment attribute of exactly one record (say corresponding to establishment e). Let E denote the set of workers employed at e in D, and E' denote the set of workers employed at e in D'. Then D and D' are neighbors if $E \subseteq E'$, and $|E| \leq |E'| \leq \max(1+\alpha)|E|, |E|+1)$*

We define our privacy definition using this notion of neighbors:

Definition 7.2 ((α, ϵ)-Employee-Employer Privacy)**.** *A randomized algorithm $\mathcal{M}$ is said to satisfy (α, ϵ)-Employee-Employer Privacy, if for every set of outputs $S \subseteq range(M)$, and every pair of strong α-Neighbors D and D', we have*

$$Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon} Pr[\mathcal{M}(D') \in S]$$

First, note that every pair of neighboring employer-employee tables must differ in the presence or absence of at least one worker. Next, note that neighboring tables do not differ in either the number of establishments or the values of their public attributes. Definition 7.1 bounds changes in a subset of the workforce of an establishment by α times the total workforce.

Theorem 7.1. *Let $\mathcal{M}$ be an algorithm satisfying (α, ϵ)-employer-employee privacy. Then, $\mathcal{M}$ satisfies the individual privacy requirement at privacy level ϵ, and the establishment size and shape requirements at privacy level (ϵ, α).*

When releasing counts over both establishment and worker attributes, this privacy notion proves too strong to provide usable results. For this purpose, we also give a weaker privacy notion. We start by defining neighbors.

[3]The number of establishments with size > 1000 is a sensitive count. This count was computed using node differential privacy with $\epsilon = 0.1$. The reported range is the 95% confidence interval based on the fact that the Laplace mechanism results in a noise of at most $\frac{1}{\epsilon} \ln \frac{1}{\delta}$ with probability $1 - \delta$.

Definition 7.3 (Weak α-Neighbors)**.** *Let D and D' be two employer-employee tables such that they differ in the employment attribute of exactly one record (say corresponding to establishment e). Let $\phi : U \to \{0, 1\}$ be any property of a worker record, and for any $E \subset U$, let $\phi(E) = \sum_{r \in E} \phi(r)$. Let E denote the set of workers employed at e in D, and E' denote the set of workers employed at e in D'. D and D' are called weak α-neighbors if for every ϕ*

$$\phi(E) \leq \phi(E') \leq \max((1+\alpha)\phi(E), 1) \tag{7}$$

Note that any property of the workforce can be represented using the function ϕ. The total employment count can be represented by the constant function that always outputs 1 for any record. The property "females with a college degree" can be represented by a ϕ that returns 1 for the records satisfying that property. Definition 7.3 bounds changes in every subset of the workforce corresponding to some of attribute values proportionally, by a factor of $(1 + \alpha)$. This neighboring definition can be used to give a weaker privacy notion.

Definition 7.4 (Weak (α, ϵ)-Employee-Employer Privacy)**.** *A randomized algorithm $\mathcal{M}$ is said to satisfy weak (α, ϵ)-Employee-Employer Privacy, if for every set of outputs $S \subseteq range(M)$, and every pair of weak α-Neighbors D and D', we have*

$$Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon} Pr[\mathcal{M}(D') \in S]$$

Theorem 7.2. *Let A be an algorithm satisfying weak (α, ϵ)-employer-employee privacy. Then, A satisfies the individual privacy requirement at privacy level ϵ and the establishment shape requirement at level (ϵ, α). A satisfies the establishment size requirement at level (ϵ, α) for weak adversaries.*

The difference between the strong and weak variants of the requirement is the following. Suppose the attacker knows there are at least Δ 19 year-old employees in an establishment, and knows the *exact counts* of employees by age for all other ages (totaling to $x - \Delta$). Thus, the attacker's only uncertainty is about the number of 19 year-olds. Then, Definition 7.2 requires that the attacker should not be able to distinguish between whether the number of 19 year-old employees is Δ or Δ' for all $\Delta \leq \Delta' \leq \Delta + \alpha \cdot x$. While one might expect few 19 year-olds in an establishment, algorithms satisfying Definition 7.2 will not be able to release that fact unless α is very small.

Under Definition 7.3 the attacker should not be able to distinguish between whether the number of 19 year-old employees is Δ or Δ' for all $\Delta \leq \Delta' \leq (1 + \alpha)\Delta$.

Keeping α fixed, larger ϵ values results in less privacy protection, since adversaries can better distinguish neighboring databases. On the other hand, when keeping ϵ fixed, larger α values result in more privacy protection.

7.2 Privacy Semantics

The bounds on an adversary's Bayes factor for the change in beliefs about the sensitive information follow from Section 4.

When $\alpha = 0$, neighboring tables differ exactly in the presence or absence of one individual worker. Thus, presence of a specific worker in the dataset is protected under (α, ϵ)-Employee-Employer Privacy, for all values of α. However, we need α to be a positive constant to ensure any privacy for establishments.

When $\alpha \to \infty$, databases are neighbors only when either E_w or E'_w is empty. This has a similar (but not identical) effect as neighbors that differ in the presence or absence of an establishment (like node differential privacy).

Next, we can also bound the extent to which an adversary can infer properties of employees and establishments beyond just neighboring databases. Both neighboring definitions, 7.1 and 7.3, induce a metric $d(\cdot)$ over possible databases.

We can use this metric to reason about what inferences an adversary can make. In general, we can show the following. For mechanism $\mathcal{M}$ satisfying one of our privacy definitions,

$$\Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon \cdot d(D,D')} \cdot \Pr[\mathcal{M}(D') \in S]. \quad (8)$$

In particular, there are two things to note. First, the distance between databases that differ in workplace attributes is infinite (we can disclose this information because the workplace attributes are public). Next, suppose D and D' are neighbors with different employment sets E_w and E'_w for establishment w, and $|E_w| = x$ and $|E'_w| = (1+\alpha)^k \cdot x$. Then, an adversary can't distinguish between D and D' based on an output with log-odds greater than $\epsilon \cdot k$; that is, $\epsilon \cdot k$ bounds the adversary's Bayes factor.

7.3 Composition

Algorithms satisfying both privacy definitions compose sequentially.

Theorem 7.3. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be (α, ϵ_1)- and (α, ϵ_2)-employer-employee private algorithms. Releasing the outputs of $\mathcal{M}_1(D)$ and $\mathcal{M}_2(D)$ results in $(\alpha, \epsilon_1 + \epsilon_2)$-employer-employee privacy. The same holds for weak (α, ϵ)-employer-employee privacy.*

Differentially private algorithms also satisfy parallel composition, which means that releasing the result of ϵ-differentially private algorithms on disjoint sets of records D_1 and D_2 also results in ϵ-differential privacy. The same is true for both Definitions 7.2 and 7.4 if the records in D_1 and D_2 pertain to distinct establishments.

Theorem 7.4. *Let D_1 and D_2 represent subsets of records from the employer-employee dataset that pertain to distinct sets of establishments. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be (α, ϵ)- and (α, ϵ)-employer-employee private algorithms. Releasing the outputs of $\mathcal{M}_1(D_1)$ and $\mathcal{M}_2(D_2)$ results in (α, ϵ)-employer-employee privacy. The same holds for weak (α, ϵ)-employer-employee privacy.*

Parallel composition is nuanced when D_1 and D_2 could pertain to distinct sets of workers from the same sets of establishments (e.g., males in New York and females in New York).

Theorem 7.5. *Let D_1 and D_2 represent subsets of records from the employer-employee dataset that pertain to distinct workers, but have records that arise from the same establishment. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be (α, ϵ)- and (α, ϵ)-employer-employee private algorithms. Releasing the outputs of $\mathcal{M}_1(D_1)$ and $\mathcal{M}_2(D_2)$ results in (α, ϵ)-employer-employee privacy. The same* does not *hold for weak (α, ϵ)-employer-employee privacy.*

8 Algorithms

We next present algorithms for answering queries under both (α, ϵ)-Employer-Employee privacy and weak (α, ϵ)-Employer-Employee privacy. Our algorithms describe how to answer a single count query (e.g., number of workers in the age 25-35 with a college degree who are employed in publicly-owned establishments in New York). This would correspond to a single cell in a marginal query (e.g., {place, ownership, age, education}). We denote these single counting queries as q_v, where $v \in dom(V)$ and q_V is the marginal query. These algorithms can be used to release all the counts in the marginal using the composition properties described in Section 7.3. Specifically, algorithms for releasing single counts can be parallel-composed under (α, ϵ)-Employer-Employee privacy for all marginals. However, algorithms for releasing single counts can be parallel-composed under weak (α, ϵ)-Employer-Employee privacy for marginals containing only establishment attributes, since such cells aggregate over distinct subsets of employers. Using algorithms for releasing a single count under weak (α, ϵ)-Employer-Employee privacy to release a marginal containing worker attributes would result in an effective privacy-loss parameter of $d \cdot \epsilon$, where d is the domain size of the worker attributes in the marginal query.

8.1 Log-Laplace Algorithm

The global sensitivity of a count query under Definitions 7.2 and 7.4 is unbounded; if the count is x, the sensitivity can be as large as αx. However, the logarithm of the count has a low global sensitivity of $\ln(1+\alpha)$. Thus the Log-Laplace mechanism (see Algorithm 1) adds Laplace noise to the log of the count.

Algorithm 1 Log-Laplace Mechanism

Require: : n : the sum of employment counts for a set of cells, α, ϵ: privacy parameters
Ensure: : $\tilde{n}$: the noisy employment count
Set $\gamma \leftarrow 1/\alpha$
$\ell \leftarrow \ln(n + \gamma)$
Sample $\eta \sim Laplace(2\ln(1+\alpha)/\epsilon)$
$\tilde{n} \leftarrow e^{\ell + \eta} - \gamma$

Theorem 8.1. *Suppose q_v is a query over only establishment attributes. Then, releasing q_v using Algorithm 1 satisfies (α, ϵ)-employer-employee privacy.*

Suppose q_v is a query over both establishment attributes and employee attributes. Then, releasing q_v using Algorithm 1 satisfies weak (α, ϵ)-employer-employee privacy.

All proofs are deferred to the Appendix. While the original Laplace mechanism is unbiased (the expectation of the noisy sum equals the true sum), the Log-Laplace mechanism is not. In particular we can show:

Lemma 8.2. *Let x denote a real number, and $\tilde{x}$ the random variable denoting the output of the Log-Laplace mechanism. Let $\lambda = 2\ln(\alpha+1)/\epsilon$. Then, when $\lambda < 1$, $E[\tilde{x}] + \gamma = (x+\gamma)/(1-\lambda^2)$. When $\lambda \geq 1$, $E[\tilde{x}]$ is unbounded.*

Theorem 8.3. *The expected squared relative error of the Log-Laplace mechanism for q_v is bounded when $\lambda = 2\ln(\alpha+1)/\epsilon$*

is less than 1*, and is given by:*

$$\mathcal{E}_{rel}(q_v) = \max_D \frac{|q_v(D) - \mathcal{M}(D)|}{q_v(D)} \leq (1+\gamma)^2 \frac{2\lambda^2 + 4\lambda^4}{(1-4\lambda^2)(1-\lambda/2)} \quad (9)$$

8.2 Smooth Sensitivity-based Algorithms

We next derive a mechanism using the smooth sensitivity framework [26]. The smooth sensitivity framework adds noise based on *local sensitivity* of the input database rather than global sensitivity across all databases. While local sensitivity can be much smaller than global sensitivity, adding noise proportional to local sensitivity does not ensure differential privacy, and hence, local sensitivity must be "smoothed."

Definition 8.1 (Local Sensitivity)**.** *Let q be a query, and $\mathcal{I}$ be a domain of datasets. The local sensitivity of query q for a dataset $x \in \mathcal{I}$ is*

$$LS_q(x) = \max_{y: y \in nbrs(x)} \|q(x) - q(y)\|_1$$

Note that the global sensitivity is the maximum local sensitivity over all databases. Nissim et al[26] show that it is sufficient to add noise proportional to any smooth function that upper bounds local sensitivity. The smallest such upper bound is called the *smooth sensitivity*.

Definition 8.2. *Let q be a query and b a smoothing parameter. Let $\mathcal{I}$ denote the universe of all datasets. The b-smooth sensitivity of query q with respect to database x is defined as*

$$S^*_{q,b}(x) = \max_j e^{-jb} A^{(j)}(x),$$
$$\text{where } A^{(j)}(x) = \max_{y \in \mathcal{I}: d(x,y) \leq j} LS_q(y),$$

and $d(x, y)$ is the smaller integer ℓ such that there exist databases $x = x_0, x_1, \ldots, x_\ell = y$, such that for all i, x_{i-1} and x_i are neighbors according to either Definition 7.1 or 7.3.

Definition 8.3 ([26])**.** *A probability distribution h is (a, b)-*admissible*, where a and b are functions of ϵ and δ, if $\forall \lambda \in \mathbb{R}, \Delta \in \mathbb{R}^d$ with $|\lambda| \leq b$ and $\|\Delta\|_1 \leq a$, and $\forall S \subseteq \mathbb{R}^d$,*

$$\Pr_{Z \sim h}[Z \in S] \leq e^{\epsilon/2} \Pr_{Z \sim h}[Z \in S + \Delta] + \frac{\delta}{2}, \text{ and} \quad (10)$$

$$\Pr_{Z \sim h}[Z \in S] \leq e^{\epsilon/2} \Pr_{Z \sim h}\left[Z \in S \cdot e^{\lambda}\right] + \frac{\delta}{2}. \quad (11)$$

We can now adapt Lemma 2.6 from [26] to show that adding noise from admissible distributions, scaled by the smooth sensitivity, generates provably private algorithms.

Theorem 8.4. *Suppose h is an (a, b)-admissible probability distribution with $\delta = 0$, and $Z \sim h$. For query q, let $S(x)$ be a b-smooth upper bound on the local sensitivity of q. Then, the algorithm $\mathcal{M}(x) = q(x) + \frac{S(x)}{a} \cdot Z$ satisfies (α, ϵ)-Employer-Employee privacy.*

We now compute the b-smooth sensitivity of our queries and describe an admissible distribution. For our problem, the local sensitivity itself is the smooth sensitivity.

Lemma 8.5. *Let q_v be a query on x. Let x_v be the maximum number of workers belonging to a single workplace and matching the conditions in v. Then, the b-smooth sensitivity of x, $S^*_{v,b}(x)$, is*

$$S^*_{v,b}(x) = \begin{cases} \max(x_v \cdot \alpha, 1) & \text{if } e^b \geq (1+\alpha), \\ \text{unbounded} & \text{otherwise}. \end{cases} \quad (12)$$

Lemma 8.6 ([26])**.** *$h(z) \propto \frac{1}{(1+|z|^\gamma)}$ is $(\epsilon/4\gamma, \epsilon/\gamma)$-admissible for $\gamma > 0$ ($\delta = 0$).*

Combining Theorem 8.4 and Lemmas 8.5 and 8.6 gives us the following algorithm. We use $\gamma = 4$ to ensure that the mean and variance of the noise distribution are bounded. Note that privacy is guaranteed only when $\alpha + 1 \leq e^{\epsilon/4}$.

Algorithm 2 Smooth Gamma

Require: : n : true count, α, ϵ: privacy parameters, $\alpha + 1 \leq e^{\epsilon/4}$

Ensure: : $\tilde{n}$: noisy count

Sample $\eta \sim \frac{1}{(1+|z|^4)}$

$\tilde{n} \leftarrow n + \frac{S^*_{v,\epsilon/4}(x)}{\epsilon/16}\eta$,

Lemma 8.7. *Suppose q_v is a query over only establishment attributes. Then releasing q_v using Algorithm 2 satisfies (α, ϵ)-Employer-Employee privacy.*

Suppose q_v is a query over both establishment and individual attributes. Then releasing q_v using Algorithm 2 satisfies weak (α, ϵ)-Employer-Employee privacy.

Lemma 8.8. *Algorithm 2 is unbiased and has expected L_1 error of $O(\frac{x_v \cdot \alpha}{\epsilon} + 1)$.*

9 Approximating Privacy

A standard relaxation of differential privacy is to allow for a small failure probability of δ that the attacker can distinguish neighboring datasets based on an output. We can similarly define $(\alpha, \epsilon, \delta)$-Employer-Employee privacy.

Definition 9.1 ((α, ϵ)-Employee-Employer Privacy)**.** *A randomized algorithm $\mathcal{M}$ is said to satisfy $(\alpha, \epsilon, \delta)$-Employee-Employer Privacy, if for every set of outputs $S \subseteq range(M)$, and every pair of strong α-Neighbors D and D', we have*

$$Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon} Pr[\mathcal{M}(D') \in S] + \delta$$

Weak $(\alpha, \epsilon, \delta)$-employer employee privacy is defined analogously.

Exact records may be released when $\delta = 1/O(n)$ (where n is the number of records). δ is the probability that a mechanism gives no privacy guarantee. Therefore if $\delta > 1/n$, a mechansim which releases the exact values for a δ fraction of the records (and 0 for all other records) satisfies this privacy definition.

δ increases rapidly with database distance. We would like to show an analogue to Equation 8 to show how δ decays with distance. That is, suppose $\mathcal{M}$ is a mechanism which satisfies $(\alpha, \epsilon, \delta)$-Employer-Employee privacy, and suppose for two

δ	α	ϵ	δ	α	ϵ
.05	.01	.105	5×10^{-4}	.01	.15
.05	.10	1.01	5×10^{-4}	.10	1.45
.05	.20	1.932	5×10^{-4}	.20	2.13

Table 2: Minimum values of ϵ given α and δ

databases D and D', $d(D, D') = d$. Then, we can only guarantee for all outputs S of $\mathcal{M}$,

$$\Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon d} \cdot \Pr[\mathcal{M}(D') \in S] + \Omega(\delta e^{(d-1)\epsilon}). \quad (13)$$

Note that for a high enough distance, the term $\Omega(\delta e^{(d-1)\epsilon})$ may become greater than one. This means that for database D' at least this far from D, we may allow an adversary to rule out without a doubt D'. This never happens under non-approximate privacy. An adversary must always have some amount of uncertainty between a pair of databases no matter how far apart.

Despite these drawbacks, allowing a small probability of failure can greatly increase the utility of an algorithm while still providing a provable guarantee. We use the framework developed in Section 8.2 to give an approximate algorithm.

Lemma 9.1 ([26]). *The Laplace distribution, $h(z) \propto \frac{1}{2} \cdot \epsilon^{-|z|}$, is $(\epsilon/2, \frac{\epsilon}{2\ln(1/\delta)})$-admissible.*

Using Theorem 8.4 and Lemma 8.5, we obtain Algorithm 3.

Algorithm 3 Smooth Laplace

Require: : n : true count, α, ϵ: privacy parameters, $\alpha + 1 \leq e^{\frac{\epsilon}{2\ln(1/\delta)}}$.
Ensure: : $\tilde{n}$: noisy count
Sample $\eta \sim Laplace(1)$
$\tilde{n} \leftarrow n + \frac{S^*_{v, \frac{\epsilon}{2\ln(1/\delta)}}(x)}{\epsilon/2}\eta,$

Lemma 9.2. *Suppose q_v is a query over only establishment attributes. Then releasing q_v using Algorithm 3 satisfies $(\alpha, \epsilon, \delta)$-employer employee privacy.*

Suppose q_v is a query over both establishment and individual attributes. Then releasing q_v using Algorithm 3 satisfies weak $(\alpha, \epsilon, \delta)$-employer employee privacy.

Lemma 9.3. *Algorithm 3 is unbiased and expected L_1 error is $O(\frac{x_v \cdot \alpha}{\epsilon} + 1)$.*

Note that the error of Algorithm 3 does *not* depend on δ. Therefore, the optimal δ for a fixed α and ϵ is the one which solves the inequality in Algorithm 3 with equality. In Table 2, we show some values of δ, ϵ, and α that work.

10 Empirical Evaluation

Dataset: The data used for these experiments were a geographic subset of the employer-employee data processed by the agency. Standard production edits and imputations had already been applied. The sample was taken from a recent snapshot for jobs held at a reference date. To be included, jobs had to qualify as "beginning-of-quarter" jobs, which means that the job (employee-employer relationship) has positive earnings in the reference quarter as well as the previous quarter. Then, the assumption is that the person was employed in the job on the first day of the reference quarter. The count of jobs in the sample was 10.9 million jobs from about 527,000 establishments.

Queries and Quality Measures: We use three types of query workloads to evaluate our algorithms.

- **Workload 1** A marginal over all establishment characteristics: industry sector, ownership, and location at the resolution of places (e.g., cities and towns).
- **Workload 2** Single queries over all establishment attributes, and over the worker attributes of sex and education.
- **Workload 3** The marginal over all establishment attributes, and sex and education.

First, we evaluate average L_1 error incurred by our algorithms for answering the above query workloads and compared it to the L_1 error incurred by the current protection system on the same workloads. We report the cost of provable privacy as a ratio between the two (error of provably private algorithms divided by the error of current SDL algorithm). Since our algorithms are randomized, error is averaged over 20 independent trials.

In addition to reporting the overall error ratio, we also compute the error ratio stratified by place-size ranges. The strata we consider are cells in the marginals with a population of 0-100, 100-10k, 10k-100k, and 100k+, respectively. We should not expect to do well in the first stratum, since one would expect a small number of small establishments in those places. Our results for Workload 3 are discussed here, but appear in the Appendix.

Next we evaluate the cost of formal privacy in ranking tasks. We ask two ranking queries:

- **Ranking 1** Rank all the cells in the marginal over industry sector, ownership, and location by total count in descending order.
- **Ranking 2** Rank all the cells in the marginal over industry sector, ownership, and location by number of employees who are female with a college degree in each cell in descending order. The results of this experiment are discussed here, and the plots appear in the appendix.

We measure error as the Spearman rank-order correlation between the ordering based on noisy counts returned by our algorithm to the ordering based on the counts output by the current SDL algorithm. As in the L_1 error case, we also report error stratified by place size.

Algorithms: We compare the Log-Laplace, Smooth Gamma and Smooth Laplace algorithms. We present results for $\epsilon \in \{0.25, 0.5, 0.67, 1.0, 2.0, 4.0\}$ and for $\alpha \in \{0.01, 0.05, 0.1, 0.15, 0.2\}$. Recall that for $\alpha = .1$, for example, an adversary should not be able to tell the difference in employments within 10% of each other. We do not plot errors for the Log-Laplace mechanism when the expectation of the noisy count is unbounded, that is when $2\ln(\alpha + 1) < \epsilon$ (see Lemma 8.2).

We do not vary δ for the Smooth Laplace algorithm as part of our evaluation, since δ does not affect the amount of noise added (and consequently does not impact the accuracy). Nevertheless, as discussed in Section 9 and Table 2, for a given α, δ imposes a lower bound on ϵ and eliminates possible choices for (α, ϵ) pairs. In our figures, we report results for pairs of (α, ϵ) that

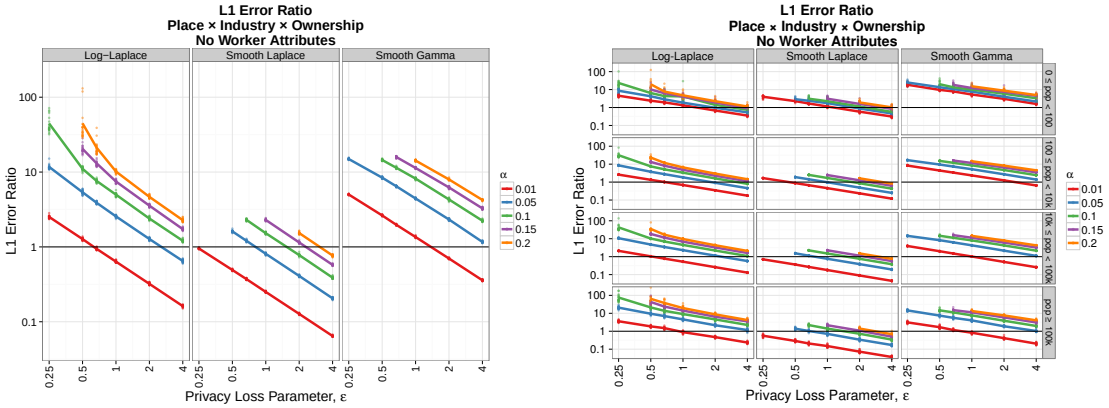

Figure 3: Average L_1 error of releasing place by industry sector by ownership marginal compared to the current system.

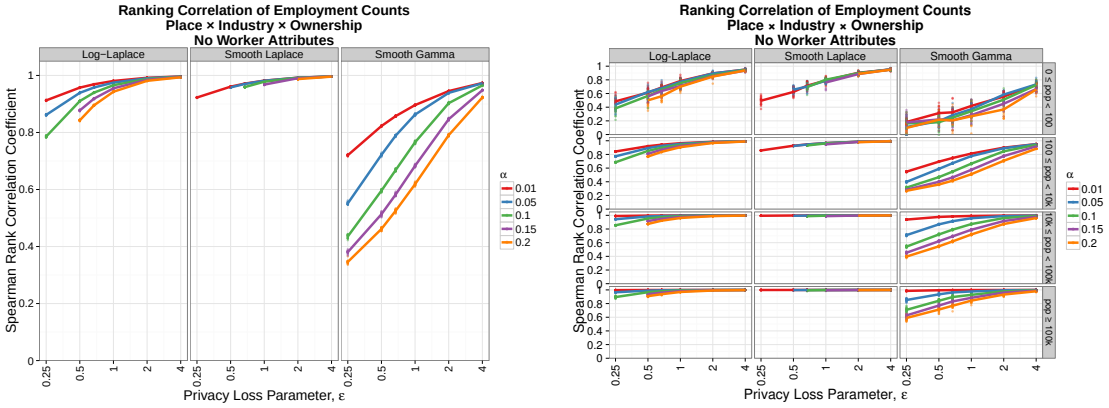

Figure 4: Spearman correlation between tested model and input noise infusion on the count of total workers ranked by place by industry section by ownership.

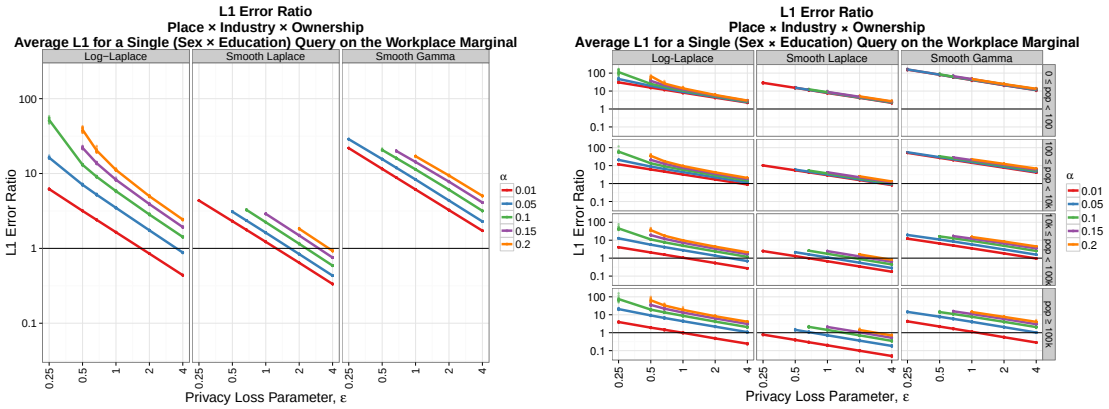

Figure 5: Average L_1 error of releasing single queries in the place by industry sector by ownership by sex by education marginal, compared to the current system.

are possible for a high failure probability of $\delta = 0.05$. The performance for smaller δ values can be read off the plots by checking whether the (α, ϵ) values are allowed for that δ.

10.1 Findings

Finding 1: For releasing marginals over only establishment attributes, our algorithms perform comparably or better than the current protection system while satisfying the strong privacy notion given in Definition 7.2. We use $\epsilon = 2$ and $\alpha = .1$ as a baseline. As can be seen in Figure 3, the average error of the Log-Laplace and Smooth Gamma algorithms is within a factor of 3 of the error of the current protection system. The Smooth Laplace algorithm performs better than the current protection system at $\epsilon = 2$ and $\alpha = .1$. Ranking results for establishment-attribute queries are shown in Figure 4. Overall, the Smooth Laplace algorithm has a correlation close to 1 when ϵ is at least 2, and the other two algorithms are close to 1 for $\epsilon \geq 4$. For rankings involving only larger population sizes, Smooth Laplace is close to 1 for all values of ϵ tested, and the Log-Laplace algorithm is close to 1 for ϵ at least 1.

We also observe that our algorithms have relative error comparable to that of the current protection system for a majority of the cells. For Log-Laplace, the relative L_1 is within 10 percentage points of the relative error of current system for 65% of the counts at $\alpha = .1$ and $\epsilon = 2$. Smooth Laplace and Smooth Gamma are within 10 percentage points for 75% and 29% of the counts, respectively.

Finding 2: For releasing individual queries over establishment and worker attributes, our algorithms perform comparably or better than the current protection system while satisfying the weaker privacy notion given in Definition 7.4. As shown in Figure 5, the average error of the Log-Laplace algorithm is within a factor of 3 of the error of the current protection system. The Smooth Laplace algorithm has nearly the same average error as the current protection system. At $\epsilon = 4$, the Smooth Laplace algorithm outperforms the current protection system for all values of α that we tested. Ranking results are shown in Figure 7. For the overall ranking, only the Smooth Laplace algorithm approaches relative error of 1 for ϵ at least 4. Restricted to larger population sizes, both Log-Laplace and Smooth Laplace perform well for all tested values of ϵ.

Finding 3: For releasing marginal queries over establishment and worker attributes, our algorithms perform worse than the current protection system, but can have acceptable performance at high values of ϵ and low values of α. As shown in Figure 6, when $\alpha \leq .05$ and $\epsilon \geq 4$, the Log-Laplace algorithm has average L_1 error within a factor of 10 of the current protection system. The Smooth Laplace algorithm is within a factor 10 for all tested values of α at $\epsilon = 4$. At the lowest tested value of α ($\alpha = .01$), the Smooth Laplace algorithm is within a factor of 3 of the error of the current system.

Finding 4: All three of our algorithms perform better as population size grows. This can be seen in both results that measure L_1 error (Figures 3, 6, and 5) as well as those that measure ranking (Figures 7 and 4). In the results that measure L_1 error, the algorithms have lower error with respect to the current protection system as the population size grows. In the ranking results, the rank order of our algorithms is more similar to the ranking order of the current protection system as the population size grows. In all cases, the improvement from the small population range (0-100) to the next smallest (100-10k) results in the largest increase in performance. Further increases in population size have a smaller effect.

Finding 5: Our Smooth Laplace algorithm performs the best of the three, and the Log-Laplace and Smooth Gamma algorithms perform similarly. Smooth Laplace performs best in all experiments, and this is not surprising since it satisfies a weaker notion of privacy. However, the ordering of the performance of the other two algorithms is not consistent. For example, in Figure 6 at lower values of ϵ, the Log-Laplace algorithm performs much worse than the Smooth Gamma algorithm. In other experiments, the Log-Laplace algorithm performs slightly better across the board. In Figure 6, the ϵ budget is divided by 10 to accommodate the lack of parallel composition in answering queries on sets of individuals belonging to the same establishment. This suggested that for smaller values of ϵ Log-Laplace does not perform as well as Smooth Gamma, but improves more rapidly as ϵ increases to eventually outperform Smooth Gamma.

Summary: Our empirical results suggest that there are a number of settings of (α, ϵ) that allow for publishing cell counts with little or no additional cost to accuracy in return for better and provable privacy protection. For marginals over only establishment attributes, all algorithms perform well at $\epsilon = 2$ and $\alpha = .1$. For individual queries of worker and establishment attributes, our algorithms perform well at $\epsilon = 2$ and $\alpha = .1$. For marginals over worker and establishment attributes, the Smooth Laplace algorithm performs well when $\epsilon \geq 4$. All algorithms perform better when the queries are over places with greater population size. Allowing for a small failure probability results in a significant reduction in error as seen with the Smooth Laplace algorithm. Counts output by our algorithms can also be used for ranking with high accuracy for $\epsilon \geq 1$.

11 Conclusions

We considered the problem of releasing employer-employee data with provable privacy guarantees and measured the utility cost of this privacy protection. We identified privacy requirements based on current interpretation of laws pertaining to the release of these data and mathematically formalized them using the Pufferfish framework. We showed that current SDL techniques do not satisfy these strong privacy requirements. Direct adaptations of ϵ-differential privacy either do not satisfy the privacy requirements, or output data with very limited utility. We develop novel privacy definitions that provably satisfy our privacy requirements. For the task of releasing marginals over establishment attributes, releasing single counts over marginals that include worker attributes, and ranking queries, our algorithms incur error that is comparable and in some cases less than the error incurred by current SDL algorithms for reasonable privacy parameters. Our results suggest that these data can be released using provably private algorithms with a low utility cost.

References

[1] J. M. Abowd, R. K. Gittings, K. L. McKinney, B. E. Stephens, L. Vilhuber, and S. Woodcock. Dynamically consistent noise infusion and partially synthetic data as confidentiality protection measures for related time-series. In *Federal Committee on Statistical Methodology Research Conference*, 2012.

[2] J. M. Abowd and I. Schmutte. Economic analysis and statistical disclosure limitation. *Brookings Papers on Econoimcs Activity*, Spring 2015.

[3] J. M. Abowd, B. E. Stephens, and L. Vilhuber. Confidentiality protection in the census bureaus quarterly workforce indicators. Technical Report TP-2006-02, U.S. Census Bureau, LEHD Program, December 2006.

[4] M. Anderson and W. Seltzer. Challenges to the confidentiality of US federal statistics, 1910-1965. *Journal of Official Statistics*, 23(1):1–34, 2007.

[5] M. Anderson and W. Seltzer. Federal statisitcal confidentiality and business data: Twentieth century challenges and continuing issues. *Journal of Privacy and Confidentiality*, 1(1):7–52, 2009.

[6] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Geo-indistinguishability: differential privacy for location-based systems. In *ACM CCS*, 2013.

[7] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, and C. Palamidessi. Broadening the scope of differential privacy using metrics. In *PETS*, 2013.

[8] S. Chen and S. Zhou. Recursive mechanism: Towards node differential privacy and unrestricted joins. In *ACM SIGMOD*, 2014.

[9] T. Dalenius. Towards a methodology for statistical disclosure control. *Statistik Tidskrift*, 15:429–444, 1977.

[10] W.-Y. Day, N. Li, and M. Lyu. Publishing graph degree distribution with node differential privacy. In *ACM SIGMOD*, 2016.

[11] G. T. Duncan, M. Elliot, and J.-J. Salazar-González. *Statistical Confidentiality Principles and Practice*. Statistics for Social and Behavioral Sciences. Springer New York, 2011.

[12] G. T. Duncan, T. B. Jabine, and V. A. de Wolf, editors. *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics; Panel on Confidentiality and Data Access [of the] Committee on National Statis- tics, Commission on Behavioral and Social Sciences and Education, National Research Council and the Social Science Research Council*. The National Academies Press, Washington, DC, 1993.

[13] C. Dwork. Differential privacy. In *ICALP (2)*, pages 1–12, 2006.

[14] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, pages 265–284, 2006.

[15] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3,4):211–407, Aug. 2014.

[16] T. Evans, L. Zayatz, and J. Slanta. Using noise for disclosure limitation of establishment tabular data. *Journal of Official Statistics*, 14(4):537–551, 1998.

[17] I. P. Fellegi. On the question of statistical confidentiality. *Journal of the American Statistical Association*, 67(337):pp. 7–18, 1972.

[18] B. A. Harris-Kojetin, W. L. Alvey, L. Carlson, S. B. Cohen, S. H. Cohen, L. H. Cox, R. E. Fay, R. Fecso, D. Fixler, G. Gates, B. Graubard, W. Iwig, A. Kennickell, N. J. Kirkendall, S. Schechter, R. R. Schmitt, M. Seastrom, M. G. Sirken, N. L. Spruill, C. Tucker, A. R. Tupek, G. D. Williamson, and R. Groves. Report on statistical disclosure limitation methodology. Research Report Statistical Policy Working Paper 22, Federal Committee on Statistical Methodology, Dec 2005.

[19] X. He, A. Machanavajjhala, and B. Ding. Blowfish privacy: Tuning privacy-utility trade-offs using policies. In *ACM SIGMOD*, pages 1447–1458, 2014.

[20] V. Karwa, S. Raskhodnikova, and A. S. G. Yaroslavtsev. Private analysis of graph structure. *PVLDB*, 4(11):1146–1157, 2011.

[21] S. Kasiviswanathan, K. Nissim, S. Raskhodnikova, and A. Smith. Analyzing graphs with node differential privacy. In *TCC*, 2013.

[22] D. Kifer and A. Machanavajjhala. No free lunch in data privacy. In *SIGMOD*, pages 193–204, 2011.

[23] D. Kifer and A. Machanavajjhala. A rigorous and customizable framework for privacy. In *PODS*, pages 77–88, 2012.

[24] D. Kifer and A. Machanavajjhala. Pufferfish: A framework for mathematical privacy definitions. *ACM Transactions on Database Systems (TODS)*, 39(1):1–36, 2014.

[25] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. In *ICDE*, page 24, 2006.

[26] K. Nissim, S. Raskhodnikova, and A. Smith. Smooth sensitivity and sampling in private data analysis. In *STOC*, pages 75–84. ACM, 2007.

[27] Y. Wang, S. Song, and K. Chaudhuri. Privacy-preserving analysis of correlated data. *CoRR*, abs/1603.03977, 2016.

A Related Work

Prior work on interpreting privacy requirements:

As reviews like [4] and the similar review for business data in [5] make clear, the privacy provisions in the statutes were designed to protect individuals and businesses from uses of the data that specifically identified them and, moreover, harmed them by subjecting them to punishment or competitive disadvantage in ways that could not be accomplished as easily without the confidential data. For protecting individuals, [17] formalized these requirements as ensuring that no exact disclosure of records in the underlying data. Fellegi derived the necessary and sufficient conditions for a set of published tables to be fully resistant to a subtraction attack that could expose one or more records. Thus, the primary goal of existing SDL techniques designed by national statistical agencies has been to prevent exact re-identification. For instance, a method called primary and complementary suppression [18] implements Fellegi's conditions [17], and has long been considered by statistical agencies around the world as compliant with confidentiality protection laws [11].

With regards to business data, protecting the characteristics of an establishment has been interpreted and implemented as described in Section 5 as ensuring that (a) the true counts of the workforce characteristics are never released or used to compute aggregates, and (b) employment counts of a workplace are perturbed by a confidential multiplicative factor unique to that workplace.

Despite avoiding exact disclosures, data publications might violate individual or business privacy by allowing too precise an inference about the true values, given the published values. This idea was first formulated in the SDL literature [9] as well as rediscovered and popularized in the computer science literature (e.g., [25, 23, 13]). We show in Section 5.2 examples of such inferences that can be made by an adversary, especially in the presence of background knowledge. Our goal is to protect employer-employee data as per the aforementioned requirements while ensuring a formal privacy notion that can limit both inferential and exact disclosures.

Customizing differential privacy: Kifer and Machanavajjhala [22] prove a no-free-lunch theorem for privacy that states that one cannot achieve privacy and utility simultaneously without making assumptions on the attacker's prior knowledge. This means that no single privacy notion (including differential privacy) ensures sufficient privacy protection for all applications and data types. Hence, recent work has focused on generalizations, both strengthening and relaxing parts of the differential privacy framework. The Pufferfish framework [23, 24], which we use in this paper, generalizes differential privacy by specifying what information should be kept secret, and the attacker's prior knowledge. He et al. [19] propose the Blowfish framework, which also generalizes differential privacy and is inspired by Pufferfish. Our privacy requirements are an instantiation of Pufferfish, and our privacy definitions can be thought of as instantiations of Blowfish. Chatzikokolakis et al [7] investigate notions of privacy that can be defined as metrics over the set of databases. These have led to the design of application specific privacy notions (e.g., [6, 27]).

Differential Privacy for complex entities: Section 6 shows that our problem can be considered as one of differential privacy on graphs, but the use of existing techniques either does not satisfy our privacy requirements or results in a complete loss of utility.

B Proofs

Claim B.1. *Differential privacy on establishments satisfies all three privacy requirements given in Section 4. Differential privacy on individuals does not satisfy all requirements.*

Proof. The proof that differential privacy on establishments satisfies the three privacy requirements is essentially the same as the proof of Theorem 7.1.

Differential privacy on individuals, however, does not satisfy the requirements. For D and D' differing in d individuals, the following holds for all outputs S:

$$\Pr[A(D) \in S] \leq e^{d\epsilon} \Pr A(D') \in S. \tag{14}$$

When an establishment e with size $|e|$ changes in size by a factor of α, the change in the number of individuals is $\alpha|e|$. As $|e|$ goes to infinity, the change in the number of individuals becomes arbitrarily large, and therefore $d\epsilon$ is not a constant. □

*of **Theorem 7.1**.* **Requirement 1:** We will denote $\Pr[\mathcal{M}(D) = w]$ as $\Pr[w]$ to make the math cleaner. We have that for some individual r,

$$\log \frac{\Pr[w \mid r \in D]}{\Pr[w \mid r \notin D]} \leq \epsilon. \tag{15}$$

Then, the individual privacy requirement of Definition 4.1 follows from Bayes' theorem.

Requirement 2: Let x and y be two numbers such that $x \leq y \leq (1+\alpha)x$. Then for all establishments e (where $|e|$ denotes the number of employees at e), we want to show that

$$\log \frac{\Pr[w \mid |e| = x]}{\Pr[w \mid |e| = y]} \leq \epsilon. \tag{16}$$

Additionally,

$$\Pr[w \mid |e| = x] = \sum_{E \subset U} \Pr[w \mid e = E] \Pr[e = E \mid |e| = x], \tag{17}$$

and

$$\Pr[w \mid |e| = y] \tag{18}$$

$$= \sum_{E' \subset U} \Pr[w \mid e = E'] \Pr[e = E' \mid |e| = y] \tag{19}$$

$$= \sum_{E \subset U : |E| = x} \sum_{E' \supset E} \Pr[w \mid e = E'] \Pr[e = E' \mid |e| = y] \tag{20}$$

$$\geq \sum_{E \subset U : |E| = x} \left[\min_{E' \supset E} \Pr[w \mid e = E'] \sum_{E' \supset E} \Pr[e = E' \mid |e| = y] \right] \tag{21}$$

$$= \sum_{E \subset U} \left[\min_{E' \supset E} \Pr[w \mid e = E'] \Pr[e = E \mid |e| = x] \right]. \tag{22}$$

Then,

$$\frac{\sum_{E \subset U} \Pr[w \mid e = E] \Pr[e = E \mid |e| = x]}{\sum_{E \subset U} [\min_{E' \supset E} \Pr[w \mid e = E'] \Pr[e = E \mid |e| = x]]} \tag{23}$$

$$\leq \max_{E \subset U} \frac{\Pr[w \mid e = E] \Pr[e = E \mid |e| = x]}{\min_{E' \supset E} \Pr[w \mid e = E'] \Pr[e = E \mid |e| = x]} \tag{24}$$

$$= \max_{E \subset U} \frac{\Pr[w \mid e = E]}{\min_{E' \supset E} \Pr[w \mid e = E']} \leq e^{\epsilon}. \tag{25}$$

Requirement 3: Requirement 3 follows from requirement 2 for a adversary. For some p, q, and z, let y be $(q-p)|e|$.

$$\frac{Pr[w \mid |e_{\mathcal{X}}|/|e| = p, |e| = z]}{Pr[w \mid |e_{\mathcal{X}}|/|e| = p, |e| = z - y]} \leq e^{\epsilon}. \tag{26}$$

Additionally,

$$\frac{Pr[w \mid |e_{\mathcal{X}}|/|e| = q, |e| = z]}{Pr[w \mid |e_{\mathcal{X}}|/|e| = p, |e| = z - y]} \leq e^{\epsilon}. \tag{27}$$

Both of these hold because the difference in the sizes of is is at most $\alpha|e|$. Therefore,

$$\frac{Pr[w \mid |e_{\mathcal{X}}|/|e| = q, |e| = z]}{Pr[w \mid |e_{\mathcal{X}}|/|e| = p, |e| = z]} \leq e^{\epsilon}. \tag{28}$$

□

*of **Theorem 7.2**.* A proof of requirement 1 is the same is in the proof of Theorem 7.1. For weak adversaries, proof of requirement 2 is also the same as Theorem 7.1. Requirement 3 follows for weak adversaries.

We can also prove requirement 3, even for strong adversaries. We want to show

$$\frac{Pr[w \mid |e_{\mathcal{X}}|/|e| = q, |e| = z]}{Pr[w \mid |e_{\mathcal{X}}|/|e| = p, |e| = z]} \leq e^{\epsilon}. \tag{29}$$

for all χ, and for $0 < p \leq q \leq \min((1+\alpha)p, 1)$. Let $x = |e|p$ and $y = |e|q$. Then we have

$$\frac{\Pr[w \mid |e_{\mathcal{X}}|/|e| = q, |e| = z]}{\Pr[w \mid |e_{\mathcal{X}}|/|e| = p, |e| = z]} = \frac{\Pr[w \mid |e_{\mathcal{X}}| = y]}{\Pr[w \mid |e_{\mathcal{X}}| = x]}. \tag{30}$$

But

$$\frac{\Pr[w \mid |e_{\mathcal{X}}| = y]}{\Pr[w \mid |e_{\mathcal{X}}| = x]} \leq e^{\epsilon}, \tag{31}$$

for all $\mathcal{X}$ when $x \leq y \leq (1+\alpha)x$ by a similar argument to the proof of Requirement 2 of Theorem 7.1.

□

*of **Theorems 7.3, 7.4, and 7.5***. The proof of sequential composition follows from Pufferfish (Theorem 9.1 of [24])

To satisfy parallel composition, we need the following: Let D_1 and D_2 be disjoint parts of the domain. Then for all databases A there exists a B such that

$$d(A \cap D_1, B \cap D_1) + d(A \cap D_2, B \cap D_2) \leq d(A, B), \tag{32}$$

where $d(\cdot)$ is the distance metric induced by our neighboring definition. In particular, we should show that if $A \cap D_1, B \cap D_1$ are neighbors, and $A \cap D_2, B \cap D_2$ are neighbors, then A, B are not neighbors. This is clearly the case when D_1 and D_2 are over different sets of establishments, since neighbors can only differ in a single establishment regardless of whether we are considering strong or weak privacy.

Next, consider some pair D_1 and D_2 over workers with shared establishments, and let e be the largest such shared establishment. Suppose in $A \cap D_1$, e has E workers, and in $B \cap D_1$, e has $(1+\alpha)E$ workers. Similarly, in $A \cap D_2$, e has E workers, and in $B \cap D_2$, e has $(1+\alpha)E$ workers. Then A and B differ in $2\alpha E$ workers, and are therefore not neighbors under out strong definition.

The same does not hold under the weak definition. □

*of **Theorem 8.1***. Consider two neighboring datasets that differ in one establishment e. S' and S be the two sets of employees of e in the two databases. Let n_- denote the sum of all other counts. We just consider the case where $S \subseteq S'$. Let $x = |S|$ and $y = |S'|$. We need to ensure privacy for two cases: (i) $x = \alpha \cdot y$, and (ii) $x = y + 1$. In case (i),

$$\begin{aligned}
\frac{P(M(D_1) = o)}{P(M(D_2) = o)} &= \frac{P(M(x + n_-) = o)}{P(M(y + n_-) = o)} \\
&= \frac{P(\eta = \ln(o+\gamma) - \ln(x + n_- + \gamma)}{P(\eta = \ln(o+\gamma) - \ln(y + n_- + \gamma)} \\
&\leq exp \ \frac{\epsilon}{\ln \alpha} \ln \ \frac{\alpha \cdot y + n_- + \gamma}{y + n_- + \gamma} \\
&\leq exp\left(\frac{\epsilon}{\ln \alpha} \cdot \ln(\alpha)\right) = e^{\epsilon}
\end{aligned}$$

In case (ii),

$$\begin{aligned}
\frac{P(M(D_1) = o)}{P(M(D_2) = o)} &= \frac{P(M(n_-) = o)}{P(M(1 + n_-) = o)} \\
&\leq exp \ \frac{\epsilon}{\ln \alpha} \ln \ \frac{1 + n_- + \gamma}{n_- + \gamma} \\
&\leq exp\left(\frac{\epsilon}{\ln \alpha} \ln(\alpha)\right) = e^{\epsilon}
\end{aligned}$$

□

*of **Theorem 8.2***.

$$E[\tilde{x}] = -\gamma + (x+\gamma) \cdot E[e^{\eta}]$$

where $\eta \sim Laplace(\lambda)$. $E[e^{\eta}]$ corresponds to the value of the moment generating function $M_{\eta}(1)$.

$$M_{\eta}(1) = E[e^{\eta}] = 1 + \sum_{n=1}^{\infty} E[\eta^n]/n!$$

Since $Laplace(\lambda)$ is an even distribution, for all odd i, $E[\eta] = 0$. Moreover, $E[\eta^{2n}] = 2n!\lambda^{2n}$. Therefore, when $\lambda < 1$

$$\begin{aligned}
E[\tilde{x}] &= -\gamma + (x+\gamma) \cdot \sum_{n=1}^{\infty} \lambda^{2n} \\
&= -\gamma + (x+\gamma)/(1-\lambda^2)
\end{aligned}$$

When λ is not bounded by 1, then the expected value is not bounded. Thus this mechanism is good only when $\lambda < 1$. □

*of **Theorem 8.3***. Let y denote $x + \gamma$, where $q_v(D) = x$ is the true sum. Similarly, let $\tilde{y}$ denote $\tilde{x} + \gamma$, where $\tilde{x}$ is the output of the multiplicative laplace mechanism. We will show that

$$E((\frac{y - \tilde{y}}{y})^2) = \frac{2\lambda^2 + 4\lambda^4}{(1-4\lambda^2)(1-\lambda/^2)}$$

The result in the theorem directly follows.

$$\begin{aligned}
E((y-\tilde{y})^2/y^2) &= E(\tilde{y}^2)/y^2 - 2E(\tilde{y})/y + 1 \\
&= E(\tilde{y}^2)/y^2 - 2/(1-\lambda^2) + 1
\end{aligned}$$

$E(\tilde{y}^2)/y^2 = E[e^{2\cdot\eta}]$, where $\eta \sim Laplace(\lambda)$. $E[e^{2\cdot\eta}]$ corresponds to the value of the moment generating function $M_{\eta}(2)$.

$$\begin{aligned}
E(\tilde{y}^2)/y^2 &= E[e^{2\cdot\eta}] = M_{\eta}(2) = 1 + \sum_{n=1}^{\infty} 2^n E[\eta^n]/n! \\
&= 1 + \sum_{n=1}^{\infty} (2\lambda)^{2n} \\
&= 1/(1-4\lambda^2) \qquad \text{when } \lambda < 1/2
\end{aligned}$$

Therefore, we have:

$$\begin{aligned}
E((y-\tilde{y})^2/y^2) &= 1/(1-4\lambda^2) - 2/(1-\lambda^2) + 1 \\
&= \frac{2\lambda^2 + 4\lambda^4}{(1-4\lambda^2)(1-\lambda/^2)}
\end{aligned}$$

□

*of **Theorem 8.4***. For $y \in nbrs(x)$, we must show that

$$\Pr[\mathcal{M}(x) \in S] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(y) \in S]. \tag{33}$$

We have

$$\begin{aligned}
\Pr[\mathcal{M}(x) \in S] &= \Pr_{Z \sim h} \ Z \in \frac{S - q(x)}{S(x)/\alpha} && (34) \\
&\leq \Pr_{Z \sim h} \ Z \in \frac{S - q(y)}{S(x)/\alpha} \ \cdot e^{\epsilon/2} + \frac{\delta}{2} && (35) \\
&\leq \Pr_{Z \sim h} \ Z \in \frac{S - q(y)}{S(y)/\alpha} \ \cdot e^{\epsilon} + \delta && (36) \\
&\leq \Pr[\mathcal{M}(y) \in S] \cdot e^{\epsilon} + \delta. && (37)
\end{aligned}$$

(35) holds by the first property of Definition 8.3. (36) holds by the second property of Definition 8.3. □

*of **Lemma 8.5***. First, we show the following general claim: For all x, let $N_1(x)$ and $N_2(x)$ be sets of neighbors of x such that $N_1(x) \cup N_2(x) = N(x)$. Then, let $LS_1(x)$ be the local sensitivity of x over $N_1(x)$ and let $LS_2(x)$ be the local sensitivity over $N_2(x)$. Then, $LS = \max(LS_1, LS_2)$. Let S_1 and S_2 be smooth upper bounds on LS_1 and LS_2. We claim that the function $S = \max(S_1, S_2)$ is a smooth upper bound on LS.

1. First, we know that S is an upper bound on both LS_1 and LS_2.

2. S is smooth because for any neighboring pair x and y, the difference between $S(x)$ and $S(y)$ is at most the difference between $S_1(x)$ and $S_1(y)$ and likewise for S_2.

We apply this to our problem by letting $N_1(x)$ be the set of neighbors of x that differ in size by exactly 1, and N_2 be the set of neighbors such that the size of an establishment's employment differs by a factor of at most ϵ. For N_1, the global and local sensitivity is 1, which is smooth. For N_2, we must give a bound on the smooth sensitivity to complete the proof.

The local sensitivity of q_v with respect to x and N_2 is the maximum amount by which any firm's (matching the criteria in v) count of employees (matching the criteria in v) can change. Note that x_v is the largest such count, and therefore

$$LS_{q_v}(x) = x_v \cdot (1+\alpha) - x_v = x_v \cdot \alpha \tag{38}$$

Then, we have

$$A^{(j)}(x) = \max_{y \in D: d(x,y) \leq j} y_v \cdot \alpha.$$

This value is maximized by maximizing y_v. The maximum value for y_v is $x_v(1+\alpha)^j$. Therefore,

$$A^{(j)}(x) = x_v \cdot \alpha(1+\alpha)^j.$$

Our smooth sensitivity is therefore

$$S^*_{v,b}(x) = \max_j (\frac{1+\alpha}{e^b})^j x_v \alpha.$$

Our databases do not have a fixed size, so j can be any positive integer, and therefore the smooth sensitivity is not necessarily bounded. When $e^b < (1+\alpha)$,

$$S^*_{v,b}(x) = \max_j (\frac{1+\alpha}{e^b})^k x_v \alpha = \lim_{j \to \infty} (\frac{1+\alpha}{e^b})^j x_v \alpha,$$

which is unbounded. When $e^b \geq (1+\alpha)$, this limit is bounded, and in this case

$$S^*_{v,b}(x) = x_v \cdot \alpha.$$

□

*of **Lemmas 8.8 and 9.3***. It is well known that the Laplace mechanism is unbiased with Laplace(λ) having mean squared error of λ^2. We prove here that $h(z) \propto \frac{1}{(1+|z|^\gamma)}$ is unbiased with bounded error for $\gamma = 4$. $h(z)$ is unbiased since

$$\mathbb{E}[h(z)] = \int \frac{z}{1+|z|^4} dz = 0.$$

The variance of $h(z)$ is given by

$$\int \frac{z^2}{1+|z|^4} dz = \frac{\pi}{\sqrt{2}} \approx 2.2.$$

The error mean squared error is therefore $(2.2)^2$. □

C Additional Figures

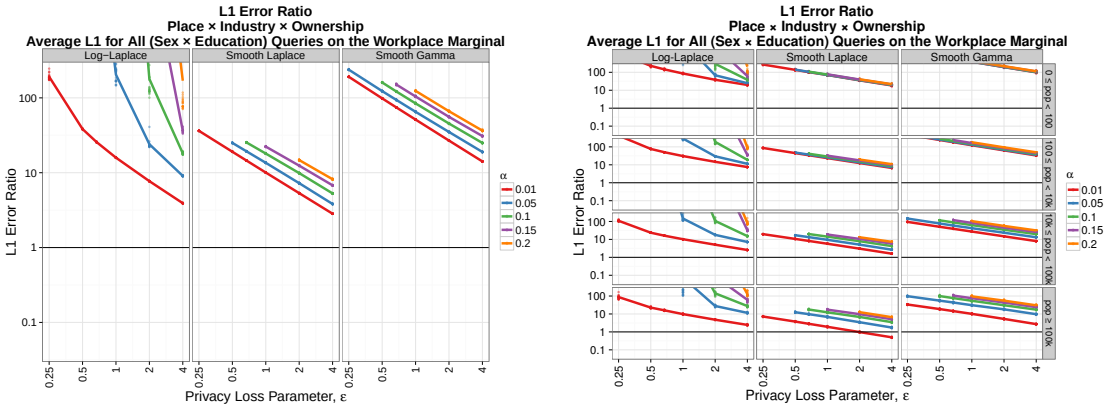

Figure 6: Average L_1 error of releasing place by industry sector by ownership by sex by education marginal compared to the current system.

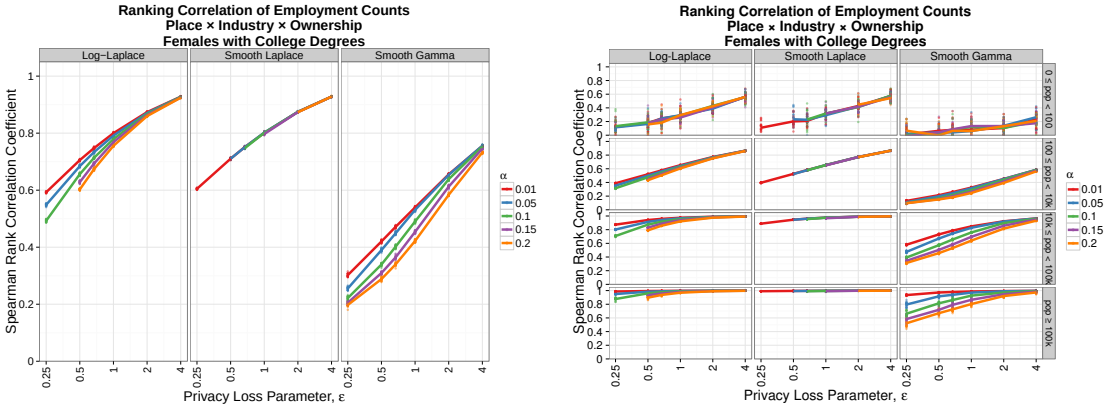

Figure 7: Spearman correlation between tested model and input noise infusion on the count of female workers who have a bachelors degree or higher, ranked by place by industry section by ownership.